1-1-2002

Geostatistical use of indirect methods for improving sampling accuracy in pastures

Alison Beth Tarr
Iowa State University

Geostatistical use of indirect methods for improving sampling accuracy in pastures

by

Alison Beth Tarr

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Crop Production and Physiology

Program of Study Committee:
Kenneth J. Moore, Major Professor
C. Lee Burras
Philip M. Dixon
Mary H. Wiedenhoeft

Iowa State University

Ames, Iowa

2002

Graduate College
Iowa State University

This is to certify that the master's thesis of

Alison Beth Tarr

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Because pasture landscapes are inherently variable, determining an approach to sampling can be difficult. When sampling is not performed adequately, the ability to characterize pasture variability is reduced. This study was performed to assess the potential of using indirect methods of soil and vegetation measurement as a means for improving sampling and mapping accuracy of a central Iowa pasture. Soil electroconductivity (EC) and canopy reflectance at eight wavebands were densely collected in a 0.42 ha nongrazed grass-legume pasture. Common directly measured soil and vegetation variables were compared with these indirect measurements on a comparatively less dense scale. A multistage sampling scheme was created based on the fuzzy clustering of soil EC and topographic characteristics. The zones based on fuzzy clustering of easily and noninvasively collected data were relatively homogeneous and an effective starting point for soil sampling of directly measured soil parameters. The geostatistical method of cokriging was coupled with the large data sets created by the indirect methods. Soil EC was examined as a covariate for improving mapping accuracy of the following target variables: soil-available P and K, pH, OM and moisture. Canopy reflectance values were used as covariates for improving mapping accuracy of above ground biomass and percent grass and legume cover. In both studies, the use of a covariate(s) generally improved local detail of the cokriged maps of soil and plant parameters. The soil EC covariate study illustrated that the larger the sampling ratio between covariate and target variable, the greater the influence on prediction of unsampled areas of soil pH and K, but not all five target variables. Prediction of unsampled soil sites using soil EC as a covariate was overall more accurate than prediction using kriging alone. When biomass and percent grass and legume cover were cokriged with the reflectance spectra most highly correlated with them, reductions in kriging variances and improved mapping accuracy were found. These results suggest that use of indirect methods such as

soil EC and canopy reflectance may save time and labor in sampling and mapping of pastures as rapidly, noninvasively and densely collected covariates.

CHAPTER 1. INTRODUCTION

Background

Pastures are often inherently variable. Developing representative sampling techniques for these heterogeneous landscapes is difficult. The common goal of sampling is to sample the minimal number of points that allows for adequate characterization of the entire pasture. The field of geostatistics can help with this dilemma. By acknowledging that points closer together in space are more similar than points farther apart, geostatistics can allow us to model spatial variation and predict values of unsampled points in a pasture. Thus, geostatistics may reduce sampling effort by capitalizing on the spatial correlation of pasture parameters of interest.

This study questions whether indirect methods can aid in the estimation of pasture parameters of interest. Use of indirect methods for agronomic measurements is an easy and rapid use of technology. By taking advantage of indirect methods, the researcher is often able to collect more data in a shorter period of time than conventional "direct" methods. As a result, data with greater resolution is created without the additional time and labor traditionally required.

In this study, apparent soil electroconductivity (EC) and multispectral canopy reflectance were utilized as indirect methods of assessing variability in soil and plant variables of interest. True to its name, the noninvasively collected data may not directly provide answers or measurements on other variables of interest. Noninvasive measurements can provide data leading us to answers, however. Field variability of the soil and above ground plant characteristics was measured using the Geonics® EM-38 and the CROPSCAN, Inc. MSR87 multispectral radiometer, respectively. From both a soil and plant sampling perspective, we are limited by the amount of information that can be collected from a given sampling scheme and density. In real life, depending on the sampling scheme and density, a sampling scheme may sample neighboring points that are separated by a bottomland area or a high pH pocket or an old building site or an old feedlot. Omitting data collection from areas such as these may lead us to incorrect field characterization and consequently, inappropriate field management. Because soil EC and spectrometer data are rapidly, easily and densely collected, it is possible that both can help to identify this in-between sampling point variability.

Using these tools, the goal was to develop more efficient sampling schemes as well as to characterize pasture variability more accurately. All data measurements were georeferenced in order to create a large spatial data set. This research included the following objectives: 1) to determine if rapidly collected, georeferenced soil information could be used to propose an accurate, multistage sampling scheme, 2) to determine if the use of soil EC as a covariate improved mapping accuracy of five soil variables across four sampling schemes and two sampling densities in a central Iowa pasture, 3) to determine the relationships between easily collected canopy reflectance data and pasture biomass and species composition and 4) to determine if the use of pasture reflectance data as a covariate improved mapping accuracy of biomass, percent grass cover and percent legume cover across three sampling schemes in a central Iowa pasture.

Thesis Organization

This thesis is organized into six distinct chapters. The first chapter includes a general introduction of methods and background for the objectives of the research. The second chapter is a literature review of the research and technical aspects pertaining to this study. Several topics are discussed at length to familiarize the scientific audience. The third through the fifth chapters are three papers written for potential publication. The third chapter involves the use of soil EC to improve sampling schemes. The fourth chapter explains the use of soil EC as a covariate in improving characterization of pasture soils. The fifth chapter contains results from the use of spectral reflectance as a covariate for improving estimation of pasture plant parameters of interest. The final chapter consists of a general summary of the research and concluding remarks regarding the potential of the methods used in this study. Lastly, acknowledgments are granted to those whom were both critical to the success of this research and influential in its progression.

CHAPTER 2. LITERATURE REVIEW

Electromagnetic Induction

Electromagnetic induction (EMI) has been used in the field of geology for many years. Subsurface electromagnetic fields have been evaluated to determine geologic structure (Geyer, 1970) as well as to locate buried metallic objects (Wait, 1971). In the last two decades, the use of EMI by soil scientists in field studies has increased. Electromagnetic induction is a noninvasive technique used to measure the electroconductivity (EC) of the soil by inducing an electrical field in the soil. The EC of a soil is determined by a combination of the water content, soluble salts, clay content and mineralogy, and soil temperature (McNeill, 1980a).

Use of EMI provides an apparent measurement of soil EC, which will be designated as EC_a. Apparent soil EC can be measured on an extremely small grid in a rapid, easy and nondestructive manner. Geonics® Ltd., Ontario, Canada manufactures instruments that measure EC_a by EMI. One of Geonics' products is the EM-38, which was used in this study. The EM-38 operates by inducing a magnetic field with a transmitting coil on one end of the instrument. The relative strength of the magnetic field varies with soil depth. The EM-38 integrates over an area approximately equal to its length of 1 m and over a depth of approximately 3 m (McNeill, 1980b). However, the measurement is primarily influenced by the 0 to 1.5 m depth increment while operated in the vertical dipole position (McNeill, 1980b). When operated in the horizontal dipole position, the primary depth of influence is the 0 to 0.75 m region (McNeill, 1980b). The receiving coil on the other end of the EM-38 reads the secondary magnetic field "induced" currents in the soil (McNeill, 1980a). Values of apparent soil EC are reported in milliSiemens per meter (mS/m).

Soil EC measured by EMI has been correlated with a variety of soil properties and crop productivity. Williams and Hoey (1987) concluded that EC_a values were highly correlated with the total clay material (<2 μm) to a depth of 15 m. Doolittle et al. (1994) found a strong inverse relationship between depth to claypan and EM-38 readings in central Missouri soils. Kachanoski et al. (1988) stated that it should be possible to predict spatial variability of soil water content using EC_a. Sheets and Hendrickx (1995) demonstrated that EMI could be used as a surrogate measure of total soil water content. Kitchen et al. (1996) mapped sand deposition using EC_a coupled with Global Positioning Systems (GPS) after the 1993 Midwest floods. Williams and Hoey (1987) also found a high correlation between EC_a and total soluble salts. In addition, corn and soybean yields have been correlated with EC_a, but the relationship was not consistent from year to year (Jaynes et al., 1995). Field measurements of EC_a identified differences in soil available N before, during, and after the corn growing season when various N treatments were applied (Eigenberg et al., 2002). Benefits from the relationship between soil EC_a and various soil and plant properties include improved mapping, prediction and management.

The use of the EM-38 in research has increased extensively because of its ability to measure data noninvasively and rapidly. When the EM-38 readings are coupled with GPS, the data becomes spatial and thus, more valuable from a mapping and spatial analysis perspective.

Multispectral Radiometry

Multispectral radiometry includes the collection of reflected energy from an object or area of interest in multiple bands of the electromagnetic spectrum (Jensen, 2000). It is based on the physical principle that molecules absorb and reflect light differently at different wavelengths. Specifically, vegetation has been shown to respond uniquely to various wavebands of light. Within the visible light region (400–700 nm), chlorophyll molecules preferentially absorb blue (400–500 nm) and red light (600–700 nm) for use in photosynthesis (Campbell, 1996). As much as 70 to 90% of incident light in these regions may be absorbed (Campbell, 1996). Peak reflectance is observed in the green region of the visible spectrum (500–600 nm); thus, the human observer sees the dominant reflection of green light as the color of living vegetation. The near infrared (NIR) region was also examined in this study. In a typical healthy green leaf, the NIR reflectance increases dramatically in the region from 700–1200 nm (Jensen, 2000). The spongy mesophyll layer in a green leaf controls the amount of NIR energy that is reflected. If plants absorbed NIR energy with the same intensity as they do in the visible region, they could become much too warm and protein denaturation could occur (Jensen, 2000).

There are more than twenty vegetation indices used for quantifying vegetative vigor (Jensen, 2000). A vegetation index is usually formed from a combination of spectral values that are added, subtracted, multiplied or divided in some manner as to form a single value associated with an area of vegetation. Birth and McVey (1968) utilized the NIR/red ratio. A general IR/red ratio also exists as a measure of vegetation vigor and abundance. The normalized difference vegetation index (NDVI) was originally proposed as a means of estimating green biomass (Tucker, 1979). It is commonly used and defined as:

$$\mathrm{NDVI} = \frac{NIR - red}{NIR + red}\,.$$

While many other indices exist, the above-mentioned indices were used in this study given the features of the available radiometer.

The radiometer used in this study measured light reflectance from the pasture canopy. Reflectance is the ratio of radiant energy reflected from the plant surface to the radiant energy incident on the plant surface. Radiometers can be found orbiting the earth on satellites (e.g. Landsat Thematic Mapper), aboard aircraft or as hand-held ground instruments. Hand-held radiometer readings can be influenced by several factors, such as incident radiation, sun angle, leaf wetness and sensor height. To minimize the effects of these factors, Guan and Nutter (2001) recommended that percentage reflectance measurements be obtained between 1100 h and 1500 h when the plant canopies are dry, with a constant sensor height and within a small range of incident radiation for all measurements.

Multispectral radiometry using hand-held radiometers has been used to estimate many plant parameters of interest. Plant greenness, biomass and yield have been correlated with reflectance measurements in various crops. Reflectance of sunlight from peanut canopies at 800 nm was positively correlated with green leaf area in peanut (Nutter, 1989; Aquino et al., 1992). Canopy light reflectance using NDVI was strongly correlated with field greenness in maize from four weeks before to four weeks after anthesis (Ma et al., 1996). Reflectance measurements were found to be successful estimators of biomass in alfalfa (Mitchell et al., 1990) and peanut (Nutter and Littrell, 1996). When NIR reflectance of a potato canopy was corrected for the soil, the NIR reflectance was useful in estimating the proportion of ground covered by potato canopies (Bouman et al., 1992). Seasonal biomass changes in tallgrass prairies were modeled using NDVI along with several other environmental variables (Olson and Cochran, 1998). Light reflectance prior to anthesis may be able to predict grain yield in corn (Ma et al., 1996), and a good correlation was found between NDVI and millet total dry matter at harvest (Lawrence et al., 2000). Furthermore, an in-season estimate of winter wheat yield was computed using the sum of two postdormancy NDVI measurements divided by the cumulative growing degree days from the first to second reading (Raun et al., 2001).

Plant disease progression has also been measured with multispectral radiometers. In peanut canopies, the reflected radiation at 800 nm provided a rapid and objective means of assessing disease severity (Nutter, 1989) as well as an evaluation of fungicide efficacy (Nutter et al., 1990). Barley stripe disease has also been highly correlated with spectral reflectance using an eight-band hand-held radiometer (Nilsson and Johnsson, 1996). In addition, herbicide injury to soybean has been evaluated using spectral reflectance data (Adcock et al., 1990).

Because greenness is an indication of plant health and adequate fertilization, spectral reflectance studies have been used to assess nitrogen fertilization of crops such as maize. Spectral differences were found among four N treatments in corn (Walburg et al., 1982). Reflectance measurements prior to anthesis (Ma et al., 1996) and at the R5 (dent) growth stage (Blackmer et al., 1996) may be useful in predicting N deficiencies.

Furthermore, the use of wavelength ratios can be used to discriminate between weed and crop species (Vrindts et al., 2002). For example, when studying weed occurrence in sugar beets, Vrindts et al. (2002) found that over 90% of the spectra studied could be classified as crop or weed. Discriminant analysis in this study also resulted in 94% correct classification of broadleaved plants in test datasets of broadleaved plants and grasses (Vrindts et al., 2002).

Spatial Data

The aforementioned tools for noninvasive data measurement were used in conjunction with GPS in this study. The addition of GPS in a data collection system allows spatial data to be created. Spatial data is data with a location. For each data value in a spatial data set, there is additional information describing the location of the data in one, two or three dimensions. Extremely large georeferenced data sets are becoming the norm with the availability of GPS equipment, high resolution imagery and high performance Geographic Information Systems (GIS).

Spatial data can tell a story. Maps of soil phosphorus concentrations can be created from spatial data (Schepers et al., 2000), location of weed species can be documented for better control (Heisel et al., 1999) and yield interpretation is made possible through collection of spatial data (Arslan and Colvin, 2002). Spatial data is often analyzed though the use of geostatistics.

Geostatistics

When doing larger-scale agricultural field research, a commonly asked question is how the property of interest varies spatially within a field. Understanding the variability within a field can help to explain processes and relationships as well as generate additional questions. Geostatistics is the field of study that allows spatial patterns to be modeled and unsampled sites to be predicted. It is useful for interpolating between measured data points to provide additional data for detailed mapping.

The development of geostatistics, also referred to as spatial statistics, occurred primarily in mining. Miners found that ore deposits in neighboring areas were more similar than areas further away. Consequently, using this relationship, miners improved their estimates for productive areas to mine. Recognizing that field variables exhibit spatial structure, Matheron reported his theory of "regionalized variables" (1971). D. G. Krige, whose name has been imprinted in geostatistics, developed and applied this theory (1966). Journel and Huijbregts provided a detailed account of the mathematics and applications of geostatistics as it related to mining (1978). From these early beginnings in mining, geostatistics has been applied to numerous other earth sciences. As applied in this thesis, geostatistics is becoming increasingly important in agronomy.

Burgess and Webster (1980a) discussed the "regionalization" of soil variables and examined various kriging models for gaining accuracy in maps of sodium content, stoniness, and cover loam in soil studies in Central Wales and Norfolk. Using the same data, Burgess and Webster (1980b) explored methods for reducing estimation variance from kriging. It was found that kriging over blocks of varying sizes rather than using point values produced smaller estimation variances. The importance of reducing estimation variances includes more accurate depictions of field variability and prediction of unsampled sites. The above methods required knowledge of the semivariogram for the soil property of interest.

Central to geostatistics is the semivariogram. It describes variation quantitatively, and it is essential for prediction of unsampled areas. Certain assumptions regarding stationarity of the variation are required to measure spatial variation in this way from a single set of observations. This is so the structure of the variation can be regarded as constant within a given region (Oliver, 1987).

For example, consider a soil property Z, such as moisture or pH that varies continuously over space. The property takes values $z(x_i)$ at location x_i, $i = 1, 2\ldots$, where x denotes a set of spatial coordinates in one, two or three dimensions. In regionalized variable theory, the quantity γ is known as the semivariance. It is the expectation of the variance at a site when sampling sites are considered pairs. Semivariances can be estimated from the following equation:

$$\gamma^*(\mathrm{h}) = \frac{1}{2N(\mathrm{h})} \sum_{i=1}^{N(\mathrm{h})} [z(\mathrm{x}_i) - z(\mathrm{x}_i + \mathrm{h})]^2$$

where N(h) is the number of pairs of values $[z(x_i), z(x_i + h)]$ separated by a lag h. The lag is a vector; thus, the estimated semivariance (γ^*) depends on both the magnitude and direction of h. [Equation taken from Vauclin et al., 1983]. A semivariogram is a graph depicting the average rate of change of a property as a function of separation distance between points. The shape of a semivariogram describes the pattern of spatial variation for a property in terms of magnitude, scale and general form (Oliver, 1987). Because the semivariogram is sensitive to outliers and to extreme values in general (Webster and Oliver, 2001), values are often transformed if the data depart from normality.

Various models exist for fitting semivariograms. Bounded models, or models where the variance reaches a maximum, include linear, circular, spherical, pentaspherical, exponential, Whittle's elementary correlation, Gaussian and pure nugget models (Webster and Oliver, 2001). The maximum variance reached in a semivariogram is called the sill variance. The distance to which pairs of points are spatially correlated is called the range. The nugget variance is the variance associated with points when h = 0, or zero separation distance. While we would expect the variance to equal zero when a sample is compared to one at the exact same location, very short range variation can occur in a property of interest. Therefore, the sampling interval may be greater than the range of spatial variation. Semivariogram models may also be combined or may fluctuate periodically (Webster and Oliver, 2001). Variation can itself vary with direction. Isotropic variation is the same in all directions from a sampling point. Anisotropic variation is dependent upon direction. It is often important to compute semivariograms for different directions (Burgess and Webster, 1980a).

Fitting of semivariogram models is often performed by least squares (Webster and Oliver, 1990). Least squares fitting assumes that the residuals from the fitted model are normally distributed and independent of one another and that estimated variances all have the same variance. It is often criticized, mainly on the grounds that the assumptions are unrealistic (Webster and Oliver, 1990). Models may also be selected on the basis of cross-validation (Vauclin et al., 1983; Warrick et al., 1986). In a cross-validation, each point in the sampling scheme is removed singly and its value is predicted based on kriging the remaining data. The resulting root mean square error (RMSE) of the cross-validation process is examined, and the variogram model with the lowest RMSE is selected (Vauclin et al., 1983; Heisel et al., 1999). Consequently, choosing models and fitting them to data remain highly controversial topics in geostatistics.

Geostatistics is based on modeling of semivariograms and prediction based on these models. The prediction of unsampled points is called kriging. The estimates are linear sums of weighted observations within a given neighborhood:

$$Z^*(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i)$$

where $Z^*(x_0)$ is the estimate of Z at x_0, N is the number of measured values involved in the estimation of the unrecorded point x_0 and λ_i are the weights associated with the ith observation (Oliver, 1987). Kriging weights depend upon the semivariogram and the configuration of the sampling points, with more weight being given to nearby points (Oliver, 1987).

Cross-correlation occurs when two or more variables are spatially correlated. Spatial relationships between topsoil silt and subsoil silt, topsoil silt and subsoil sand and subsoil silt and sand have been described by McBratney and Webster (1983), between available water content and sand content (Vauclin et al., 1983), between zinc content and elevation data (Leenaers et al., 1990), and between soil salinity and apparent soil electroconductivity (Triantafilis et al., 2001). The cross-correlation between two variables can be modeled by estimating a cross semivariogram, or cross-variogram. The cross-variogram can be estimated by the following equation:

$$\gamma_{12}^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z_1(x_i) - z_1(x_i + h)]\ [z_2(x_i) - z_2(x_i + h)]$$

where $N(h)$ is the number of pairs of values $\{[z_1(x_i), z_1(x_i + h)], [z_2(x_i), z_2(x_i + h)]\}$ separated by a lag h. Values for both variables (z_1 and z_2) must be available at the same point in order for the cross-variogram to be calculated. [Equation taken from Vauclin et al., 1983].

If one of the two variables is more difficult or more expensive to measure, then the cross-correlation between variables can be capitalized upon. For instance, if z_2 is more difficult to measure than z_1, then it may be possible to estimate values of z_2 using the values of both z_1 and z_2 and the cross-correlation between them. This method is known as cokriging. Estimating the value of z_2 can be performed with the following calculation:

$$Z_2^*(x_0) = \sum_{i=1}^{N_1} \lambda_{1i} Z_1(x_{1i}) + \sum_{j=1}^{N_2} \lambda_{2j} Z_2(x_{2j})$$

where λ_{1i} and λ_{2j} are the weights associated with Z_1 and Z_2, and N_1 and N_2 are the number of neighbors of Z_1 and Z_2 involved in the estimation of point x_0, respectively. [Equation taken from Vauclin et al., 1983]. Thus, the value of the more difficult (and thus, often undersampled) variable, Z_2 is computed by a weighted average of the observed values of the covariate (Z_1) and Z_2 occurring in the estimation neighborhood of each kriging location.

Cokriging has been implemented to estimate values of an undersampled target variable using information from its spatial correlation with a covariate, or subsidiary variable. For example, topsoil silt has been cokriged with subsoil silt and sand as covariates (McBratney and Webster, 1983), $NaHCO_3$-extractable P cokriged with 25% HCl-extractable P (Trangmar et al., 1986), NO_3 cokriged with soil EC (Zhang et al., 1992), soil EC of soil paste extract cokriged with apparent soil EC (Vaughan et al., 1995) and *Lamium* spp. weed distribution cokriged with silt content (Heisel et al., 1999). These agronomic examples include use of both invasively and noninvasively collected covariates. The added bonus of a noninvasively collected covariate is its ability to be collected in a more rapid and easier manner than invasive soil parameters such as sampling for soil nutrient and pH determination.

Geostatistics and Sampling

Geostatistics can also be used to develop a sampling strategy for a field. Because computation of the semivariogram generates estimation variances, variances can be found for any sampling density and separation distance. This would be valuable to an agriculturist who is trying to design an optimal sampling scheme with as few points as possible and with as little error as possible. For example, Burgess et al. (1981) studied sampling strategies in which the sampling effort was as little as possible for a desired precision level. Intuitively, the farther an interpolated point is from actual sampling points, the greater the error associated with the prediction of the interpolated point. Consequently, when designing a sampling grid where the variation is known to exist isotropically, Burgess et al. (1981) concluded that an equilateral triangular grid was optimal because the maximum distance between an interpolated point and its nearest sampling neighbor is minimized. However, a square grid at the same sampling density was nearly as good and would likely be preferred for convenience in layout and sampling. Where variation was anisotropic, a rectangular grid with observations made by sampling parallel transects in the direction of maximum variance was suggested (Burgess et al., 1981). In addition, with a specified level of error, the optimal sampling interval can be determined, as shown by Burgess et al. (1981). McBratney et al. (1981) defined a "recipe for an optimal sampling strategy," which included choosing the maximum estimation variance tolerable, sampling transects in three or more directions with randomly located starting points in the field, calculating the experimental semivariograms, fitting a model and determining grid spacing for a direction of maximum variance in the field based on the initial error variance allowed.

References

Adcock, T.E., F.W. Nutter, Jr., and P.A. Banks. 1990. Measuring herbicide injury to soybeans (Glycine max) using a radiometer. Weed Science 38:625-627.

Aquino, V.M., F.M. Shokes, R.D. Berger, D.W. Gorbet, and T.A. Kucharek. 1992. Relationships among late leafspot, healthy leaf area duration, canopy reflectance, and pod yield of peanut. Phytopathology 82:546-552.

Arslan, S., and T.S. Colvin. 2002. Grain yield mapping: yield sensing, yield reconstruction, and errors. Precision Agriculture 3:135-154.

Birth, G.S., and G.R. McVey. 1968. Measuring the color of growing turf with a reflectance spectrophotometer. Agron. J. 60:640-643.

Blackmer, T.M., J.S. Schepers, G.E. Varvel, and E.A. Walter-Shea. 1996. Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies. Agron. J. 88:1-5.

Bouman, B.A.M., D. Uenk, and A.J. Haverkort. 1992. The estimation of ground cover of potato by reflectance measurements. Potato Research 35:111-125.

Burgess, T.M., and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil Sci. 31:315-331.

Burgess, T.M., and R. Webster. 1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. J. Soil Sci. 31:333-341.

Burgess, T.M., R. Webster, and A.B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. Journal of Soil Science 32:643-659.

Campbell, J.B. 1996. Introduction to Remote Sensing. Second ed. Guilford Press, New York.

Doolittle, J.A., K.A. Sudduth, N.R. Kitchen, and S.J. Indorante. 1994. Estimating depths to claypans using electromagnetic induction methods. J. Soil and Water Cons. 49:572-575.

Eigenberg, R.A., J.W. Doran, J.A. Nienaber, R.B. Ferguson, and B.L. Woodbury. 2002. Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. Agriculture, Ecosystems and Environment 88:183-193.

Geyer, R.G. 1970. Effect of subsurface geologic structure in electromagnetic induction prospecting. The Mines Magazine 60(6):7-8.

Guan, J., and F.W. Nutter, Jr. 2001. Factors affecting the quality and quantity of sunlight reflected from alfalfa canopies. Plant Dis. 85:865-874.

Heisel, T., A.K. Ersboll, and C. Andreasen. 1999. Weed mapping with co-kriging using soil properties. Precision Agriculture 1:39-52.

Jaynes, D.B., T.S. Colvin, and J. Ambuel. 1995. Yield mapping by electromagnetic induction. in Site -Specific Management for Agricultural Systems. ASA-CSSA-SSSA, Madison, WI.

Jensen, J.R. 2000. Remote sensing of the environment: an earth resource perspective. Prentice Hall, Upper Saddle River, New Jersey.

Journel, A.G., and C.J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York.

Kachanoski, R.G., E.G. Gregorich, and I.J. Van Wesenbeeck. 1988. Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods. Can. J. Soil Sci. 68:715-722.

Kitchen, N.R., K.A. Sudduth, and S.T. Drummond. 1996. Mapping of sand deposition from 1993 midwest floods with electromagnetic induction measurements. J. Soil and Water Cons. 51:336-340.

Krige, D.G. 1966. Two-dimensional weighted moving average trend surfaces for ore-evaluation. Journal of the South African Institute of Mining and Metallurgy 66:13-38.

Lawrence, P.R., B. Gerard, C. Moreau, F. Lheriteau, and A. Buerkert. 2000. Design and testing of a global positioning system-based radiometer for precision mapping of pearl millet total dry matter in the Sahel. Agron. J. 92:1086-1095.

Leenaers, H., J.P. Okx, and P.A. Burrough. 1990. Employing elevation data for efficient mapping of soil pollution on floodplains. Soil Use and Management 6:105-114.

Ma, B.L., M.J. Morrison, and L.M. Dwyer. 1996. Canopy light reflectance and field greenness to assess nitrogen fertilization and yield of maize. Agron. J. 88:915-920.

Matheron, G. 1971. The theory of regionalized variables and its applications, Vol. Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau, No. 5. Ecole Nationale Superieure des Mines de Paris.

McBratney, A.B., and R. Webster. 1983. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. Journal of Soil Science 34:137-162.

McBratney, A.B., R. Webster, and T.M. Burgess. 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables. I. Theory and method. Computers & Geosciences 7:331-334.

McNeill, J.D. 1980a. Electrical conductivity of soils and rocks. Technical Note TN-5. Geonics, Ltd., Mississauga, Ontario, Canada.

McNeill, J.D. 1980b. Electrical conductivity of soils and rocks. Technical Note TN-6. Geonics, Ltd., Mississauga, Ontario, Canada.

Mitchell, A.R., P.J. Pinter, Jr., J.N. Guerrero, C.B. Hernandez, and V.L. Marble. 1990. Spectral reflectance measurements of alfalfa under sheep grazing. Agron. J. 82:1098-1103.

Nilsson, H., and L. Johnsson. 1996. Hand-held radiometry of barley infected by barley stripe disease in a field experiment. Journal of Plant Diseases and Protection 103:517-526.

Nutter, F.W., Jr. 1989. Detection and measurement of plant disease gradients in peanut with a multispectral radiometer. Phytopathology 79:958-963.

Nutter, F.W., Jr., and R.H. Littrell. 1996. Relationships between defoliation, canopy reflectance and pod yield in the peanut-late leafspot pathosystem. Crop Protection 15:135-142.

Nutter, F.W., Jr., R.H. Littrell, and T.B. Brenneman. 1990. Utilization of a multispectral radiometer to evaluate fungicide efficacy to control late leaf spot in peanut. Phytopathology 80:102-108.

Oliver, M.A. 1987. Geostatistics and its application to soil science. Soil Use and Management 3:8-20.

Olson, K.C., and R.C. Cochran. 1998. Radiometry for predicting tallgrass prairie biomass using regression and neural models. J. Range Management 51:186-192.

Raun, W.R., J.B. Solie, G.V. Johnson, M.L. Stone, E.V. Lukina, W.E. Thomason, and J.S. Schepers. 2001. In-season prediction of potential grain yield in winter wheat using canopy reflectance. Agron. J. 93:131-138.

Schepers, J.S., M.R. Schlemmer, and R.B. Ferguson. 2000. Site-specific considerations for managing phosphorus. Journal of Environmental Quality 29:125-130.

Sheets, K.R., and J.M.H. Hendrickx. 1995. Noninvasive soil water content measurement using electromagnetic induction. Water Resources Research 31:2401-2409.

Trangmar, B.B., R.S. Yost, and G. Uehara. 1986. Spatial dependence and interpolation of soil properties in West Sumatra, Indonesia: II. Co-regionalization and co-kriging. Soil Sci. Soc. Am. J. 50:1396-1400.

Triantafilis, J., I.O.A. Odeh, and A.B. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci. Soc. Am. J. 65:869-878.

Tucker, C.J. 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing Environment 8:127-150.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175-184.

Vaughan, P.J., S.M. Lesch, D.L. Corwin, and D.G. Cone. 1995. Water content effect on soil salinity prediction: a geostatistical study using cokriging. Soil Sci. Soc. Am. J. 59:1146-1156.

Vrindts, E., J. De Baerdemaeker, and H. Ramon. 2002. Weed detection using canopy reflection. Precision Agriculture 3:63-80.

Wait, J.R. 1971. Electromagnetic induction technique for locating a buried source. IEEE transactions on geoscience electronics 9:95-98.

Walburg, G., M.E. Bauer, C.S.T. Daughtry, and T.L. Housley. 1982. Effects of nitrogen nutrition on the growth, yield, and reflectance characteristics of corn canopies. Agron. J. 74:677-683.

Warrick, A.W., D.E. Myers, and D.R. Nielsen. 1986. Geostatistical methods applied to soil science, *In* A. Klute, ed. Methods of soil analysis, Part 1. 2nd ed. Agronomy 9:53-82.

Webster, R., and M.A. Oliver. 1990. Statistical methods in soil and land resource survey. Oxford University Press, New York.

Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists. John Wiley & Sons, Ltd., Chichester.

Williams, B.G., and D. Hoey. 1987. The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. Australian Journal of Soil Research 25:21-27.

Zhang, R., D.E. Myers, and A.W. Warrick. 1992. Estimation of the spatial distribution of soil chemicals using pseudo-cross-variograms. Soil Sci. Soc. Am. J. 56:1444-1452.

CHAPTER 3. USE OF SOIL ELECTROCONDUCTIVITY IN A MULTISTAGE SAMPLING SCHEME

Introduction

Devising a sampling scheme in a pasture situation is difficult due to the inherent variability in pasture landscapes. The design of traditional sampling schemes such as grid and triangular overlook a very important truth in field studies: certain areas of a field are more similar than other areas of the field. This truth is also the fundamental principle of geostatistics: points that are located close together are often more similar than points located far apart. This principle can be applied to generating a sampling scheme for a field situation that is more efficient.

A "good" sampling scheme should be able to quantify field variability as accurately as possible and with as few points as possible. Cline (1944) quantified the number of sampling points required to achieve a specified level of error in soil measurements. Using the classical formulae for estimating means and variances and for choosing the size of a sample, *n*, Cline gave the equation:

$$n = t_{\alpha}^{2}s^{2}/(x-\mu)^{2}$$

where s^2 is the estimated variance, t_α is the value of Student's t at the chosen level of probability, α, and $(x-\mu)$ represents the tolerable deviation between the true mean μ and the sample x in the estimation of the mean. This use of classical statistics fails to recognize the spatial correlation among soil variables and can often overestimate the required number of samples. Recognizing that field variables exhibit spatial structure, Matheron reported his theory of "regionalized variables" (1971). Burgess and Webster discussed the "regionalization" of soil variables and examined various kriging methods for gaining accuracy in maps of soil variables as well as precision in sampling strategy in soil survey

(1980a; 1980b; Burgess et al., 1981). Thus, the once asked question of "how many samples" now incorporates the question "where to sample".

An initial step in sampling a field may be to divide the field into a number of homogeneous strata (Cline, 1944). Stratified sampling requires taking one or multiple samples within each stratum (Sampford, 1962). If the area within the strata were homogeneous, this would support the principle of geostatistics by grouping together points within a field that are similar. Two-stage sampling designs begin with an initial sampling of primary units and then secondary units are selected from each of the selected primary units (Sampford, 1962; Thompson, 1992). Further stages of sampling from the secondary or higher-order units may follow and are termed multistage sampling. The question now becomes, how does one know if areas of a field are similar if samples have not yet been taken?

One way to quickly measure field variability is by use of electromagnetic induction (EMI). Soil electrical conductivity (EC) can be measured on an extremely small grid in a rapid, easy and nondestructive manner by the use of EMI. Soil EC is a measurement that is affected primarily by a combination of soil water content, dissolved salt content, clay content and mineralogy and soil temperature (McNeill, 1980a). Soil EC measured by EMI has been correlated with clay content (Williams and Hoey, 1987; Doolittle et al., 1994), soil water content (Kachanoski et al., 1988; Sheets and Hendrickx, 1995), sand deposition (Kitchen et al., 1996), total soluble salts (Williams and Hoey, 1987), yield (Jaynes et al., 1995), and soil available N (Eigenberg et al., 2002). Benefits from the relationship between soil EC and various soil properties include improved soil mapping, prediction and management.

In this study, the coupling of EMI with global positioning systems (GPS) was capitalized on to provide rapidly, easily, and nondestructively collected georeferenced data. By using this information to measure field variability within the pasture, a sampling scheme was devised. It makes sense to sample a field more densely in areas that are heterogeneous and less densely in areas that are homogeneous; thus, a stratified sampling scheme was used. The objective of this study was to determine if rapidly collected, georeferenced soil information could be used to propose an accurate, multistage sampling scheme.

Materials and Methods

Research was conducted at the Iowa State University Rhodes Research Farm (41°52′N, 93°10′W) in central Iowa. The Wisconsin loess-covered landscape has an underlying Yarmouth-Sangamon paleosol. The soils are primarily slope and erosion phases of the Fayette and Clarinda series (Fenton, 2001). The pasture site of study included topographically distinct summit, sideslope, toeslope, backslope and opposite summit landscape positions (Fig. 1).

A dense sampling grid consisting of 116 points was devised for a 0.42 ha, nongrazed grass-legume pasture (Fig. 2). Sampling points were arranged in a triangular grid with inter- and intra-row separation distances of 6 m. In order to obtain data from samples located closer than 6 m, an additional point was sampled within each row at randomly chosen 1 or 2 m separation distances. This short range variation in soil samples was investigated in order to obtain a more reliable experimental semivariogram model (Burgess and Webster, 1980a; Kravchenko and Bullock, 2002).

Apparent soil EC was measured with the Geonics® EM-38 as it was pulled behind a four-wheel drive vehicle in a nonconductive cart (Fig. 3). The EM-38 was operated in the vertical dipole orientation at 0.20 m above the soil surface. The EM-38 integrates over an area approximately equal to its length of 1 m and over a depth of approximately 3 m (McNeill, 1980b). However, the measurement is primarily influenced by the 0 to 1.5 m depth increment (McNeill, 1980b). The vertical rather than horizontal dipole orientation was used so that any fluctuations in carrying height of the EM-38 above the soil surface would have little impact on EC readings (McNeill, 1980b). Given the speed of travel and rate of EC

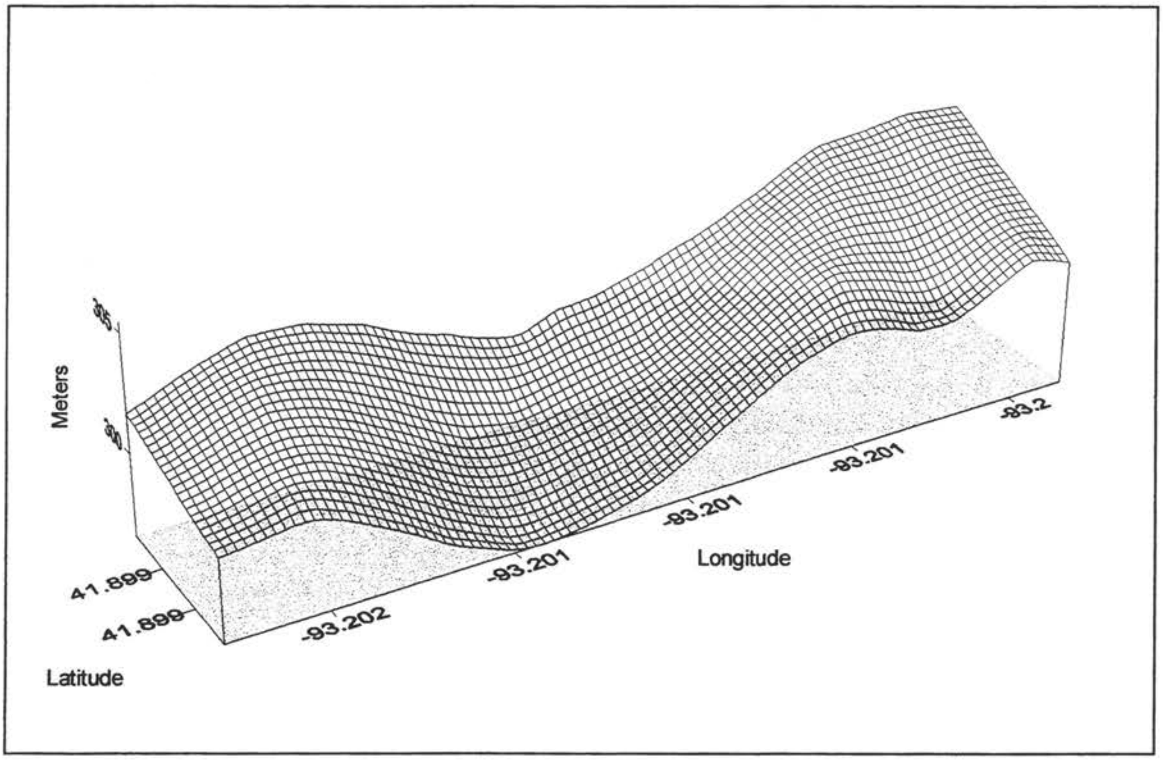

Figure 1. Wireframe map illustrating topographic variation of pasture site (Golden Software, Inc., 1999). Meters shown above mean sea level.

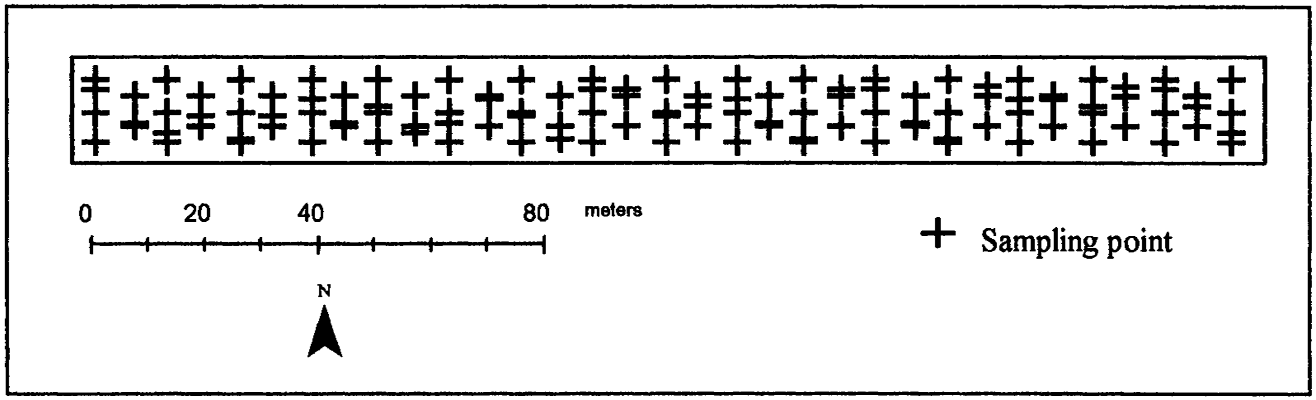

Figure 2. Initial dense sampling scheme (n = 116).

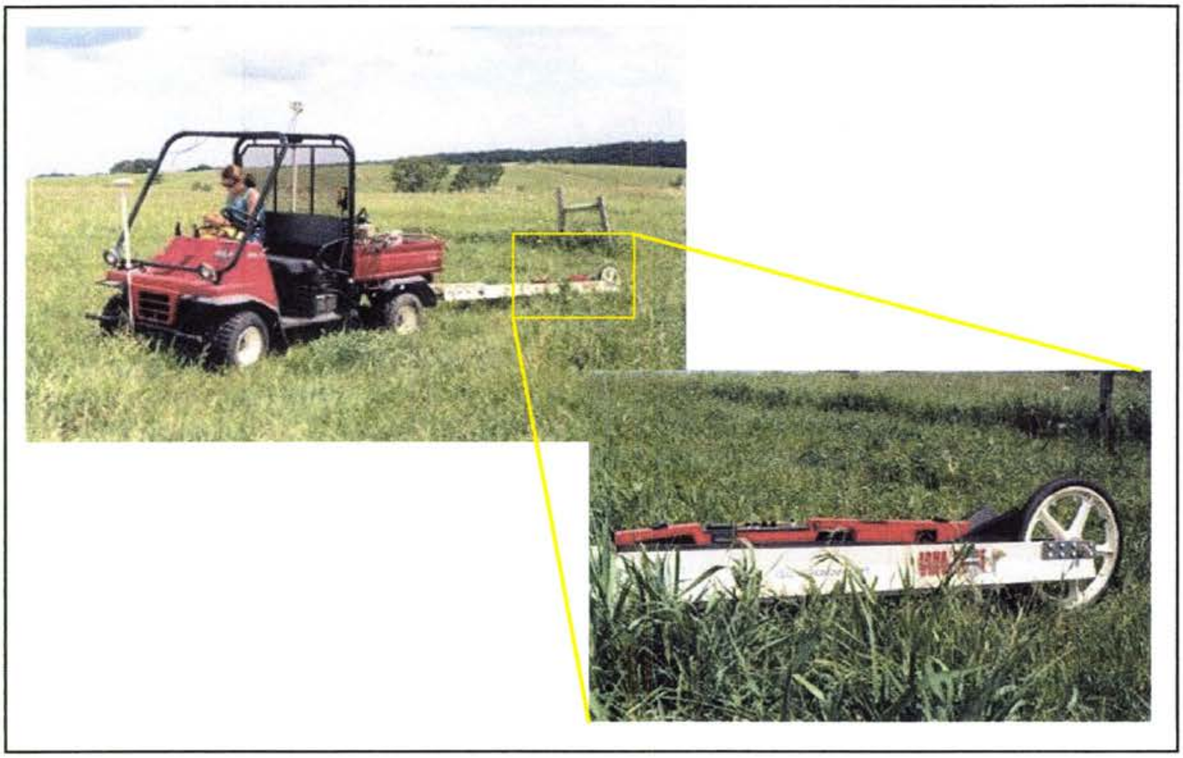

Figure 3. Geonics® EM-38 placed in nonconductive cart with GPS receiver mounted on four-wheel drive vehicle. Inset is a close-up of EM-38.

data collection, an EC reading was logged approximately every 4 m in 2 m transects throughout the pasture. Each soil EC measurement was georeferenced using a Trimble GPS Pathfinder® Pro XR receiver, and all GPS locations were differentially corrected (DGPS) to obtain 1-2 m accuracy. EM-38 measurements were recorded via a direct connection to Trimble System Controller (TSC) 1 Asset Surveyor software in the GPS datalogger. Positional data for the soil EC values were corrected for the lag distance between GPS receiver and the EM-38 instrument. Soil EC values at each of the 116 grid points were interpolated from the dense data set of the 834 georeferenced EC points.

Soil samples were collected via coring following the EM-38 measurements. At each of the 116 sampling sites, five 15 cm soil cores were collected and combined for soil pH and available P and K analysis in the Iowa State University Soils Testing Laboratory (Ames, IA). Organic matter was analyzed from a single core composed of three depth increments to 51 cm, using a dry combustion method. Average organic matter percentage over all three depth increments was reported. To quantify soil moisture, a single core composed of seven 15 cm samples was taken. Each 15 cm sample was analyzed using the gravimetric moisture method (Buckman and Brady, 1971), and an average percent soil moisture was reported.

Elevation data were recorded using a Corvallis Microtechnology, Inc. (CMT) real time kinematic (RTK) system, and slope data were calculated from this using ArcView 3.2 Spatial Analyst (ESRI, 1996). Geostatistical analyses were performed using ArcView 8.1 ArcGIS Geostatistical Analyst (ESRI, 2001).

Multistage Sampling

The fuzzy k-means algorithm was implemented as the first step in the multistage sampling scheme. The algorithm was used to stratify the field into relatively homogeneous zones based on densely collected soil parameters. The fuzzy k-means method has been utilized for classifying soil and landscape data when binary or strictly discrete groupings are not adequate to describe natural systems (Burrough, 1989; McBratney and de Gruijter, 1992; Odeh et al., 1992; Irvin et al., 1997). In addition, fuzzy classification has been practiced with yield data and remotely sensed imagery (Lark, 1998; Boydell and McBratney, 2002). Given the continuous nature of soils, fuzzy set classification provides a suitable means of classifying areas of a field. A combination of topographic attributes and apparent soil EC were used to delineate zones using the software program Management Zone Analyst© (University of Missouri-Columbia and Agricultural Research Service, 2000). The clustering algorithm was iterated for 3-9 zones. Based on the fuzziness performance index (FPI) and normalized classification entropy (NCE), five zones appeared optimal for establishing strata of homogeneity in the pasture. The FPI and NCE performance indices were discussed and used to evaluate the varying number of climatic classes by McBratney and Moore (1985). Within these five strata, two intensities of a ranked set sampling scheme (McIntyre, 1952) were analyzed: n = 30 and n = 15 (Figs. 4 and 5). Ranked set sampling was first described by McIntyre (1952) as a method for obtaining more precise and unbiased measurements of forage yield. The five strata delineated by clustering corresponded to the sets in ranked set sampling. The points within each of the five strata were ranked in order of EC value magnitude. Soil EC was chosen as a concomitant variable because it is the variable most easily and accurately collected (Patil et al., 1994), and it was correlated with the soil variables of interest. When constructing the n = 30 scheme, six points were selected in each of the five strata. With this scheme, the six points were chosen based on maximizing the within-zone variation of soil EC. The six points were selected based on choosing the minimum, maximum, and four in-between quantile values of soil EC within each zone. Maximum within-zone variation was sought in order to maintain the variability identified throughout the field. Similarly, when devising the n = 15 scheme, three points were selected in each of the five strata with the goal of maximizing the within-zone variation of the

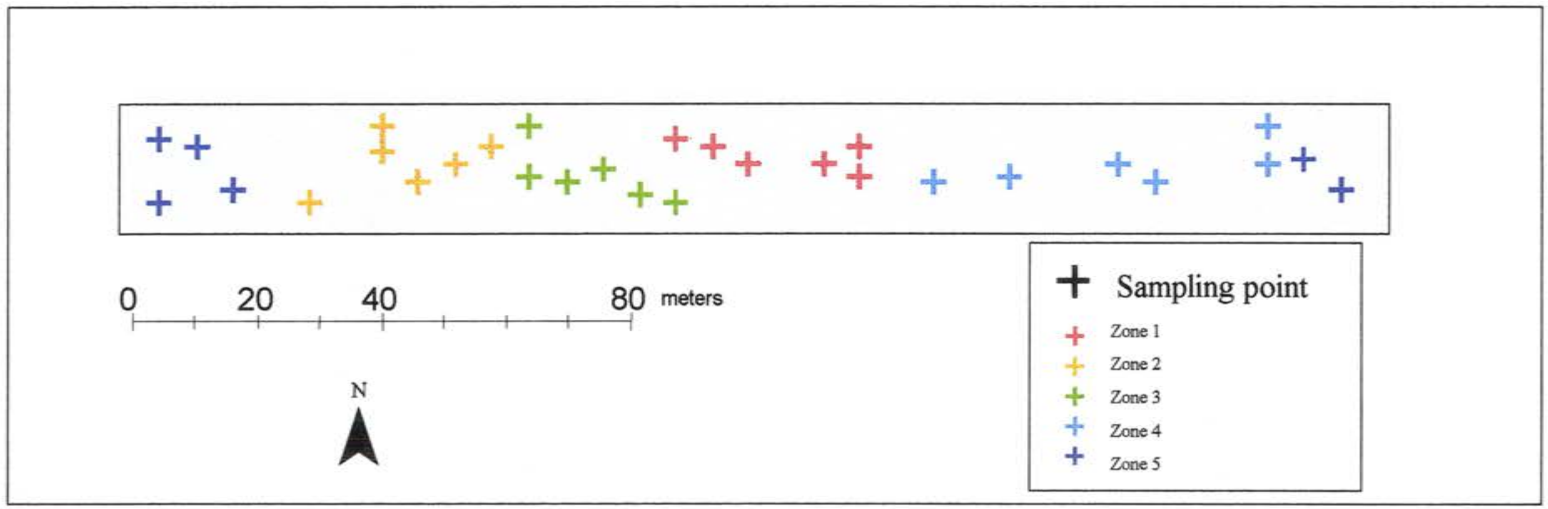

Figure 4. Sampling points for multistage sampling scheme (n = 30).

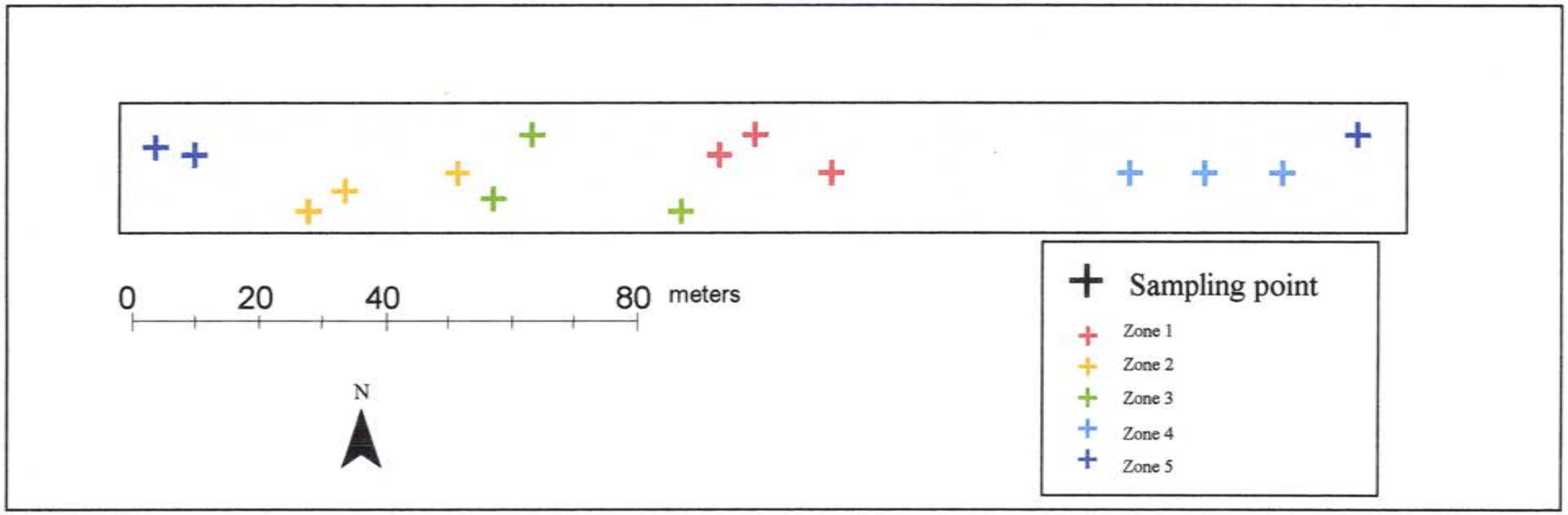

Figure 5. Sampling points for multistage sampling scheme (n = 15).

concomitant variable, soil EC. The three points were selected based on choosing the minimum, maximum and median values of soil EC within each zone.

Results and Discussion

Statistical Data Analysis

Statistical summaries of topographic and soil attributes for the initial, dense grid (n = 116) sampling scheme are presented in Table 1. With a range in elevation of 6.4 m, a CV of 0.7 did not describe the actual variability in topography. The CV was very small due to the relatively large mean value for elevation as compared to the SD. Large CVs and ranges for soil P, K, pH, OM and moisture illustrated the magnitude of soil variability within the pasture (Table 1). The degree to which soil EC can detect this soil variability improves our ability to delineate homogeneous sampling strata.

How well soil EC identifies soil variability depends upon soil EC's relationship with the soil parameters of interest. The relationships among the chemical and physical attributes for the initial, dense sampling scheme are shown in Table 2. The results of the large database indicate a strong correlation between soil EC and soil pH, elevation and soil K (r values greater than 0.5). Soil EC was moderately correlated with soil moisture and OM values ($0.10 < r < 0.50$) and weakly correlated with soil P and slope ($r \leq |0.10|$) (Table 2).

The relationships from Table 2 were not of primary interest in our objectives, however. Of interest was determining whether soil EC could be used effectively to identify soil spatial variation and zones of homogeneity from which to sample. In the pasture of study, variability in soil EC appeared closely related to the variation in landscape position and depth to paleosol (Fenton, 2002). Higher values of soil EC were measured in the toeslope positions. These positions have a higher moisture content and are underlain by a clay-textured paleosol (Clarinda series). Lower values of soil EC were measured upslope on the summit positions where the soils were developed entirely in loess. Middle values were

Table 1. Descriptive statistics of measured topographic and soil attributes for 116 sampling points.

	Depth	Mean	SD	CV	Min.	Max.
Elevation, m	-	301.2	2.1	0.7	296.6	303.0
Slope, degrees	-	5.4	1.6	29.6	1.7	8.8
P-Bray, ppm†	0.15 m avg	14.9	7.1	47.7	4.0	51.5
K-NH_4AcO, ppm	0.15 m avg	159.8	59.1	37.0	56.0	369.0
pH, 1:1 soil/water	0.15 m avg	6.0	0.4	7.3	5.4	7.3
OM, %	0.51 m avg	2.2	0.8	36.8	1.1	6.1
Moisture, %	1.07 m avg	27.6	2.3	8.4	24.2	41.8
Soil EC, mS/m	≅1.5 m	44.3	4.7	10.5	35.5	55.0

† ppm is parts per million, OM is organic matter, mS/m is milliSiemens per meter.

Table 2. Partial correlation matrix among chemical and physical soil attributes for 116 sampling points.

	Elevation	Slope	P-Bray	K	pH	OM	Moisture	Soil EC
Elevation	1							
Slope	-0.02	1						
P-Bray	-0.15	-0.27	1					
K-NH_4AcO	0.65	0.08	0.14	1				
pH	-0.75	-0.31	0.14	-0.67	1			
OM	-0.42	-0.21	0.44	-0.26	0.41	1		
Moisture	-0.46	-0.15	0.31	-0.30	0.51	0.40	1	
Soil EC	-0.62	-0.10	-0.06	-0.56	0.70	0.24	0.17	1

measured on the backslope where the soils are formed in loess and the underlying paleosol. Because the soil was not dominated by carbonates, the variation in soil EC values was likely related to soil moisture content and textural properties (Brevik and Fenton, 2002). Textural properties influence soil parameters such as organic matter and ionic properties; thus, it was concluded that soil EC is measuring variation in soil properties related to moisture and texture.

Performance of Fuzzy Classification

Using elevation, slope and apparent soil EC data, the fuzzy k-means clustering algorithm resulted in the delineation of 5 zones throughout the pasture (Figure 6). This clustering agreed well with landscape position. Zone 1 included bottomland and backslope characteristics; Zone 2 was primarily a sideslope; Zone 3 was primarily bottomland; Zone 4 combined all three landscape positions, but with more gently rolling sideslopes compared to Zone 2; Zone 5 was a region consisting primarily of summit land but with some sideslope area. Cluster membership was spatially discrete with only a few points that were nonadjacent to other members of their zone. Zone 5 included points on both the west and east ends of the pasture, but this separation in zone membership was likely due to the repeating landscape pattern in the field. Zone 5 revealed a repeated summit from the repeating summit, sideslope, toeslope and backslope pattern in the field. Table 3 quantitatively describes each of the five strata. It is worth noting that the five zones are not of equal size. The fuzzy clustering algorithm reiteratively classified each of the 116 points until each point in a zone was more similar to the cluster centroid than to any other cluster. Therefore, a relatively small amount of variability in soil measurements was expected within each zone. Although Zone 3 was primarily a bottomland area, it exhibited the most variability in soil P, pH, organic matter, moisture and EC. Zone 5 showed the most variability in soil K.

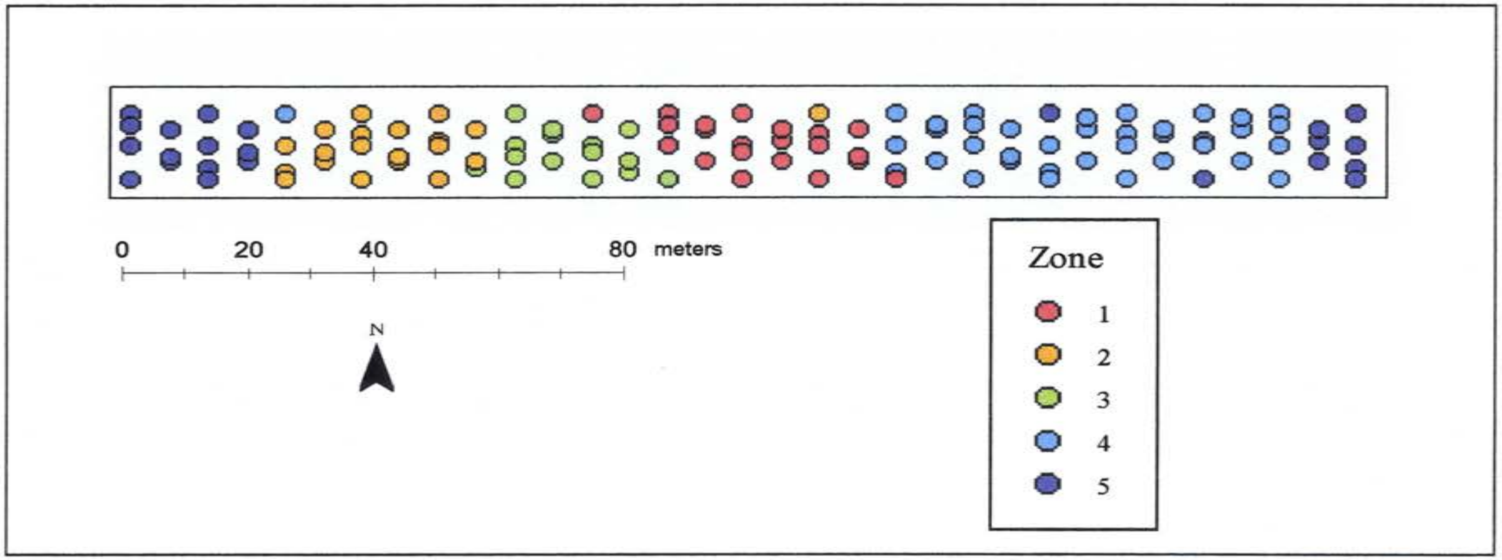

Figure 6. Fuzzy clustering results for the initial sampling scheme (n = 116).

Table 3. Mean and standard deviation for chemical and physical soil attributes shown for each zone.

	Zone 1		Zone 2		Zone 3		Zone 4		Zone 5	
	n = 21		n = 20		n = 15		n = 37		n = 23	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Elevation, m	298.6	1.0	299.0	1.0	297.0	0.2	301.7	0.5	301.5	1.0
Slope, degrees	5.5	1.0	7.6	0.8	4.0	1.0	5.9	0.7	3.6	0.7
P-Bray, ppm†	12.6	7.3	12.9	3.3	21.9	12.0	13.4	4.4	15.8	5.6
K-NH_4AcO, ppm	111.4	30.3	153.1	30.6	100.8	34.7	193.9	49.1	192.3	64.8
pH, 1:1 soil/water	6.4	0.3	5.8	0.1	6.7	0.4	5.7	0.1	5.8	0.1
OM, %	2.4	1.0	2.1	0.6	3.1	1.2	2.0	0.5	2.0	0.5
Moisture, %	27.6	1.7	27.1	0.8	31.1	4.6	27.2	0.9	26.7	0.8
Soil EC, mS/m	52.1	1.8	42.1	1.9	47.6	3.5	42.1	2.0	40.7	1.6

† ppm is parts per million, OM is organic matter, mS/m is milliSiemens per meter.

With the five strata delineated, an analysis of variance using SAS (SAS Institute, Inc., 2001) was performed to determine if the strata were indeed different based upon the measured soil variables. Results from the ANOVAs for soil P, K, pH, OM and gravimetric moisture are shown in Table 4 based on the entire data set (n = 116). The results indicated that the effect of zone on all the soil parameters was significant. Thus, the resultant zones from the fuzzy clustering algorithm appear to have delineated significantly different zones based on the five soil attributes.

Identifying whether or not the zones were unique clusters was important. If the zones were significantly different from one another based on the soil parameters, then they would be considered acceptable sampling zones. A multiple comparison test was run to determine the pairwise differences among zones for the five soil attributes (SAS Institute, Inc., 2001). Results from the multiple comparison tests are reported for both the noninvasively measured variables that were used for the fuzzy clustering algorithm (Table 5) and for the invasively sampled soil parameters (Table 6). As demonstrated, not all of the zones were significantly different from each other for the soil parameters. When comparing the *t*-tests of the zones for the noninvasive soil parameters to the invasive soil parameters, it was evident that the zones appear much more unique with the noninvasive soil parameters (i.e. there were more P values < 0.20) (Tables 5 and 6). This result was logical because the clustering algorithm was run with these variables and mathematically, the zones would be different. The zones from the invasively measured soil variables were not as distinct from each other based on the *t*-tests (Table 6). This result was not surprising because all the soil variables did not vary spatially in the same manner that soil EC, elevation and slope did. For example, Zones 4 and 5 differed significantly in soil EC and slope, but they were only significantly different in one of the five measured soil variables ($P < 0.20$). This result indicated Zones 4 and 5 were more similar physically and chemically than suggested by fuzzy clustering based on noninvasively collected soil parameters. The remainder of the pairwise comparisons demonstrated the

Table 4. One-way ANOVA results for five zones.

Source	df	P-Bray		$K\text{-}NH_4AcO$		pH		OM		Moisture	
		Mean Square	*F* ratio	Mean Square	*F* ratio	Mean Square	*F* ratio	Mean Square	*F* ratio	Mean Square	*F* ratio
Zone	4	254.34	6.00***	42406.78	20.27***	3.92	70.17***	4.01	7.38***	52.22	14.31***
Error	111	42.36		2091.75		0.06		0.54		3.65	

*** Significant at the 0.001 probability level.

Table 5. Pairwise *t*-tests comparing zones for the soil variables used for fuzzy clustering.

Zone comparison	Soil variables used for fuzzy clustering					
	Soil EC		Slope		Elevation	
	Diff. between means	$P > \|t\|$	Diff. between means	$P > \|t\|$	Diff. between means	$P > \|t\|$
1 – 2	10.0	<.0001	-2.0	<.0001	-0.4	0.11
1 – 3	4.5	<.0001	1.5	<.0001	1.6	<.0001
1 – 4	10.0	<.0001	-0.4	0.11	-3.1	<.0001
1 – 5	11.4	<.0001	2.0	<.0001	-2.9	<.0001
2 – 3	-5.5	<.0001	3.6	<.0001	2.0	<.0001
2 – 4	-0.04	0.95	1.7	<.0001	-2.7	<.0001
2 – 5	1.3	0.04	4.0	<.0001	-2.5	<.0001
3 – 4	5.4	<.0001	-1.9	<.0001	-4.7	<.0001
3 – 5	6.8	<.0001	0.4	0.12	-4.5	<.0001
4 – 5	1.4	0.02	2.3	<.0001	0.2	0.28

Table 6. Pairwise t-tests comparing zones for the five measured soil variables.

Zone comparison	Measured soil variables									
	Phosphorus		Potassium		pH		OM		Moisture	
	Diff. between means	$P > \|t\|$	Diff. between means	$P > \|t\|$	Diff. between means	$P > \|t\|$	Diff. between means	$P > \|t\|$	Diff. between means	$P > \|t\|$
1 – 2	-0.3	0.90	-41.7	0.004	0.6	<.0001	0.3	0.15	0.5	0.45
1 – 3	-9.2	<.0001	10.6	0.50	-0.3	0.0001	-0.7	0.006	-3.5	<.0001
1 – 4	-0.7	0.67	-82.6	<.0001	0.7	<.0001	0.4	0.03	0.4	0.44
1 – 5	-3.2	0.11	-80.9	<.0001	0.6	<.0001	0.4	0.08	0.9	0.14
2 – 3	-9.0	0.0001	52.3	0.001	-0.9	<.0001	-1.0	<.0001	-3.9	<.0001
2 – 4	-0.5	0.79	-40.8	0.001	0.1	0.11	0.1	0.62	-0.04	0.93
2 – 5	-2.9	0.14	-39.2	0.006	0.06	0.40	0.07	0.76	0.4	0.48
3 – 4	8.5	<.0001	-93.1	<.0001	1.0	<.0001	1.1	<.0001	3.9	<.0001
3 – 5	6.0	0.006	-91.4	<.0001	0.9	<.0001	1.1	<.0001	4.3	<.0001
4 – 5	-2.5	0.16	1.7	0.89	-0.04	0.48	-0.03	0.86	0.5	0.37

significant differences between zones. Seven of the ten pairwise comparisons were significantly different ($P < 0.20$) for soil P, eight of ten for soil K, eight of ten for soil pH, seven of ten for soil OM and five of ten for soil moisture (Table 6). These significant differences communicate confidence in the delineation of sampling zones.

Performance of Multistage Sampling Scheme

Fuzzy k-means classification techniques can result in a reduction in sampling intensity because homogeneous areas are not oversampled (Odeh et al., 1990). The new stratified sampling scheme derived from the fuzzy k-means algorithm decreased the number of sampling points from 116 to 30 and 116 to 15, which resulted in decreases of 74 and 87%, respectively. A sampling scheme consisting of fewer sampling points while maintaining accuracy of field characterization was the goal. One method of assessing the benefit of the new multistage sampling scheme is to compare its variance of the estimated population total to that of a random sample of the same size (Sampford, 1962). When comparing the $n = 30$ stratified sampling scheme with that of a random sample of 30 points, a general trend of increasing variances was observed with the random sample for four of the five soil variables (Table 7). Potassium was the single exception for the $n = 30$ sampling scheme. This difference in variances was not significant ($P = 0.20$), however, and it may be attributed to the inherently high variability in soil K for this pasture (Table 3). The same general trend of increasing variances with random samples was also evident in the $n = 15$ sampling scheme. The single exception was soil P, and again, this difference occurred with a relatively variable soil property (Table 3) and it was not significant ($P = 0.20$). Although few of the stratified-random comparisons were significantly different from one another ($P = 0.20$), the difference in population variances may transfer to a measurable difference in actual field characterization. This effect was not examined. However, reducing the estimation of the population variance improves the precision associated with the sample values.

Table 7. Comparison of estimated population variances between the stratified and random sampling schemes at two sampling densities (n = 30 and n = 15).

	Estimated population total variances							
	n = 30				n = 15			
	Stratified	Random	*F* ratio	*P* > *F*	Stratified	Random	*F* ratio	*P* > *F*
Phosphorus	8803	11554	1.31	0.232	18244	16518	0.91	0.571
Potassium	909319	870687	0.96	0.544	2459667	2667591	1.08	0.442
pH	14	54	3.86	<0.001	48	102	2.13	0.077
OM	99	144	1.45	0.157	250	267	1.07	0.449
Moisture	1586	2460	1.55	0.118	924	1071	1.16	0.389

It is worthwhile to note that the stratified sampling scheme significantly reduced population variance the most for soil pH in both the n = 30 and n = 15 sampling schemes (P = 0.10) (Table 7). It is hypothesized that this is because pH is the soil variable most closely correlated with the three noninvasively measured soil parameters used for fuzzy classification (Table 2). Consequently, as would be expected, stratification worked best for soil variables most closely related to the variable(s) used for fuzzy classification.

Validation of the two new sampling schemes was conducted by interpolating the data from the new sampling schemes to estimate unsampled points in the pasture. Data were interpolated by kriging (Burgess and Webster, 1980a). The predictions resulting from kriging were compared to the actual values from the unsampled points known from the original (n = 116) sampling scheme (Fig. 2). Similar to most soil sampling situations, we assumed that we did not know the semivariogram model for the soil variables measured. Therefore, when attempting to find the best semivariogram model, the model with the lowest root mean square error (RMSE) of prediction for cross-validation of the sampled points was chosen (Triantafilis et al., 2001). The RMSE was a measure reporting the precision of prediction. It should be as small as possible for unbiased and precise predictions (Triantafilis et al., 2001). For cross-validation, each sampling point from a sampling scheme was removed in turn and kriging was used to predict its value based on the remaining points. The ArcGIS extension Geostatistical Analyst was used for these procedures (ESRI, 2001).

The 86 and 101 points not included in the n = 30 and n = 15 sampling schemes, respectively, were used as validation sets. These large validation sets were compared with the predicted soil values from kriging (Table 8). A comparison of predicted versus actual values for the soil parameters described how well the kriged results from each sampling scheme estimated unsampled points. The pH at the unsampled points was predicted the best out of the five soil variables (Table 8). This result could again be attributed to the fact that pH is the soil variable most closely correlated with the three noninvasively measured soil

Table 8. Validation set root mean square errors (RMSE) of soil data prediction for kriging the n = 30 and n = 15 sampling schemes. Coefficient of determination (r^2) between predicted and actual values for validation set.

	n = 30		n = 15	
	RMSE	r^2	RMSE	r^2
	-- unit --		-- unit --	
P-Bray, ppm†	4.903	0.549	6.884	0.265
K-NH_4AcO, ppm	45.70	0.518	44.160	0.458
pH, 1:1 soil/water	0.169	0.858	0.306	0.555
OM, %	0.762	0.187	0.733	0.250
Moisture, %	1.611	0.501	2.414	0.044

† ppm is parts per million, OM is organic matter.

parameters used for fuzzy classification (Table 2). The results of soil pH indicated that the higher sampling density resulted in a nearly twofold decrease in prediction error, and it also improved the r^2 value (Table 8). The increase in prediction error associated with the decrease in sampling intensity may or may not be of consequence depending on the cost to benefit ratio of additional sampling. In the case of soil pH, the additional cost and effort required to sample the 15 additional points would have to be weighed against the potential benefits from more precise application of lime.

The results of soil phosphorus and moisture were similar to soil pH (Table 8). There was an increase in prediction error for the unsampled points when fewer samples were measured, and the coefficient of determination was significantly smaller for the lower density sampling scheme. Soil moisture exhibited an especially poor r^2 value for the $n = 15$ sampling scheme. Several hypotheses may account for this result. First, this result may be due to the fact that soil moisture gradients did not correlate well with the zones produced by fuzzy classification of the three noninvasively measured soil parameters (Table 2). Secondly, the selection of points within in each zone may not have been ideal for accurate detection of pasture variability in moisture. Thirdly, the reliability of the experimental semivariogram is affected by the size of the sample and the configuration of the sample (Webster and Oliver, 2001). With only 15 sampling points and an irregular sampling pattern, the resultant experimental semivariogram may not have provided an accurate model of soil variability. In fact, Webster and Oliver (2001) state that experimental variograms based on fewer than 50 data often have little or no evident structure. However, as the size of the sample is increased, the structure of the variogram becomes clearer (Webster and Oliver, 2001). In the field, however, the size of the sample is often determined by the availability of resources such as time, labor and money.

Soil potassium and organic matter displayed different results than pH, phosphorus and moisture. With both potassium and organic matter, the RMSE of prediction decreased when fewer points were sampled (Table 8). However, this decrease in RMSE was not substantial. The r^2 value for potassium decreased when fewer points were sampled (similar to pH, phosphorus and moisture). The r^2 value for organic matter actually increased for the n = 15 sampling scheme. This aberrant outcome was likely a rare result.

Conclusions

Soil EC can be used to detect soil variability in pastures. Using EMI techniques, one can take advantage of a rapid, easy and noninvasive method of data collection. When soil EC data was coupled with georeferenced topographic information such as elevation and slope, a large database was available to describe field variability. This large database was incorporated with fuzzy clustering as a way to delineate homogeneous sampling zones in the pasture. While these zones were not unique clusters for every soil parameter measured, they did provide an effective starting point for soil sampling (Tables 5 and 6). As a final stage in the multistage sampling scheme, ranked set sampling insured an unbiased selection of points (McIntyre, 1952) while maintaining within cluster variability.

The study suggests that soil variables most closely related to those used for clustering are predicted with the least error at unsampled points. Soil pH was most highly correlated with soil EC (Table 2), and the prediction accuracy of pH at unsampled points was highest (Table 8). In general, there was a loss in prediction accuracy resulting from a decrease in sampling intensity (McBratney and Webster, 1983). However, this loss in predictive accuracy may not have economic or management consequence to the producer.

Stratification was useful in dividing a heterogeneous population such as a pasture into relatively homogeneous subpopulations, or sampling zones. By stratification of this pasture using the fuzzy k-means clustering algorithm, a more directed approach to soil sampling was taken. A more optimal sampling scheme covers the same area with fewer sampling points, less time and less labor while using rapid, noninvasive, geospatial tools. Knowing about field variation without an invasive and time-consuming survey may save labor and can direct efforts for sampling.

References

Boydell, B., and A.B. McBratney. 2002. Identifying potential within-field management zones from cotton-yield estimates. Precision Agriculture 3:9-23.

Brevik, E.C., and T.E. Fenton. 2002. Influence of soil water content, clay, temperature, and carbonate minerals on electrical conductivity readings taken with an EM-38. Soil Surv. Horiz. 43:9-13.

Buckman, H.O., and N.C. Brady. 1971. The nature and properties of soils. 7th ed. Macmillan, New York.

Burgess, T.M., and R. Webster. 1980a. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil Sci. 31:315-331.

Burgess, T.M., and R. Webster. 1980b. Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging. J. Soil Sci. 31:333-341.

Burgess, T.M., R. Webster, and A.B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. Journal of Soil Science 32:643-659.

Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40:477-492.

Cline, M.G. 1944. Principles of soil sampling. Soil Science 58:275-288.

Doolittle, J.A., K.A. Sudduth, N.R. Kitchen, and S.J. Indorante. 1994. Estimating depths to claypans using electromagnetic induction methods. J. Soil and Water Cons. 49:572-575.

Eigenberg, R.A., J.W. Doran, J.A. Nienaber, R.B. Ferguson, and B.L. Woodbury. 2002. Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. Agriculture, Ecosystems and Environment 88:183-193.

ESRI. 1996. ArcView Spatial Analyst, Redlands, CA.

ESRI. 2001. ArcView 8.1. Release ArcGIS Geostatistical Analyst, Redlands, CA.

Fenton, T.E. 2001. Personal communication. Department of Agronomy, Iowa State University, Ames, Iowa.

Fenton, T.E. 2002. Personal communication. Department of Agronomy, Iowa State University, Ames, Iowa.

Golden Software, Inc. 1999. Surfer Version 7.0, Golden, Colorado.

Irvin, B.J., S.J. Ventura, and B.K. Slater. 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77:137-154.

Jaynes, D.B., T.S. Colvin, and J. Ambuel. 1995. Yield mapping by electromagnetic induction. in Site -Specific Management for Agricultural Systems. ASA-CSSA-SSSA, Madison, WI.

Kachanoski, R.G., E.G. Gregorich, and I.J. Van Wesenbeeck. 1988. Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods. Can. J. Soil Sci. 68:715-722.

Kitchen, N.R., K.A. Sudduth, and S.T. Drummond. 1996. Mapping of sand deposition from 1993 midwest floods with electromagnetic induction measurements. J. Soil and Water Cons. 51:336-340.

Kravchenko, A.N., and D.G. Bullock. 2002. Spatial variability of soybean quality data as a function of field topography: I. Spatial Data Analysis. Crop Science 42:804-815.

Lark, R.M. 1998. Forming spatially coherent regions by classification of multi-variate data: an example from the analysis of maps of crop yield. International Journal of Geographical Information Science 12:83-98.

Matheron, G. 1971. The theory of regionalized variables and its applications, Vol. Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau, No. 5. Ecole Nationale Superieure des Mines de Paris.

McBratney, A.B., and R. Webster. 1983. How many observations are needed for regional estimation of soil properties? Soil Science 135:177-183.

McBratney, A.B., and A.W. Moore. 1985. Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology 35:165-185.

McBratney, A.B., and J.J. de Gruijter. 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. Journal of Soil Science 43:159-175.

McIntyre, G.A. 1952. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research 3:385-390.

McNeill, J.D. 1980a. Electrical conductivity of soils and rocks. Technical Note TN-5. Geonics, Ltd., Mississauga, Ontario, Canada.

McNeill, J.D. 1980b. Electrical conductivity of soils and rocks. Technical Note TN-6. Geonics, Ltd., Mississauga, Ontario, Canada.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1990. Design of optimal sample spacings for mapping soil using fuzzy-k-means and regionalized variable theory. Geoderma 47:93-122.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy-c-means: Application to classification and soil-landform interrelationships. Soil Sci. Soc. Am. J. 56:505-516.

Patil, G.P., A.K. Sinha, and C. Taillie. 1994. Ranked set sampling, p. 167-200, *In* G. P. Patil and C. R. Rao, eds. Environmental Statistics, Handbook of Statistics, Vol. 12. Elsevier Science, Amsterdam.

Sampford, M.R. 1962. An introduction to sampling theory with applications to agriculture. First ed. Oliver and Boyd, Edinburgh.

SAS Institute, Inc. 2001. SAS Version 8.2. Release 8.2, Cary, NC.

Sheets, K.R., and J.M.H. Hendrickx. 1995. Noninvasive soil water content measurement using electromagnetic induction. Water Resources Research 31:2401-2409.

Thompson, S.K. 1992. Sampling. John Wiley & Sons, Inc., New York.

Triantafilis, J., I.O.A. Odeh, and A.B. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci. Soc. Am. J. 65:869-878.

University of Missouri-Columbia and Agricultural Research Service. 2000. Management Zone Analyst software. Release 1.0.

Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists. John Wiley & Sons, Ltd., Chichester.

Williams, B.G., and D. Hoey. 1987. The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. Australian Journal of Soil Research 25:21-27.

CHAPTER 4. IMPROVING MAP ACCURACY OF SOIL VARIABLES USING SOIL ELECTROCONDUCTIVITY AS A COVARIATE

Introduction

Interest in management of field variability has increased with the increasing availability and adoption of precision agriculture tools and technology. Because of this increased interest, methods for better spatial characterization of a field are valuable. Spatial characterization may include the collection and mapping of such variables as plant health, field productivity and soil nutrients. Remote sensing imagery can identify differences in vegetation greenness at pixel sizes of 10 m from thousands of kilometers away (Jensen, 1996), and georeferenced yield monitors allow measurement of productivity by the second. While these tools allow increased spatial information to be known about above ground productivity at a high resolution, the basis for most soil management decisions, soil tests, are measured at a relatively low resolution. Because of the high cost, soil tests are measured at a much lower density when compared to the sensors available for greenness measurements and productivity. Gaining more information from traditional soil sampling would aid in the creation of more accurate soil maps and thus, provide direction for more precise management of field inputs such as nutrients and lime.

The obvious way to gain more information from soil sampling is to take more samples. This option can be both cost- and labor-prohibitive, however. Another option is to find a cheaper and easier method that can measure variability of a secondary soil parameter that may help to explain the variability of the soil parameter(s) of interest. This relationship can subsequently be used to improve the prediction variance associated with the soil variable of interest. For example, Vauclin et al. (1983) incorporated secondary information on the spatial variability of sand content to help predict values of two primary variables of interest, available water content and water stored at 1/3 bar. McBratney and Webster (1983a) estimated topsoil silt content utilizing its spatial interrelationship with subsoil sand and subsoil silt as secondary variables. Trangmar et al. (1986) explored the use of densely sampled 25% HCl-extractable P as a covariate for estimating the variable of interest, $NaHCO_3$-extractable P. Noninvasively collected measures have also served as favorable secondary variables. Leenaers et al. (1990) improved the quality of interpolation maps of zinc concentration in the Geul floodplain in the southern Netherlands by incorporating elevation data. Jaynes (1996) found that electromagnetic induction (EMI) conductivity data reduced the estimation error of soil organic carbon fraction by 33% compared to kriging only soil organic carbon samples.

All of these examples incorporate a geostatistical method called cokriging. Geostatistics is a field of study that involves modeling and prediction of spatial variability (Journel and Huijbregts, 1978). Kriging is a method of interpolation used when a variable displays spatial dependence. Cokriging is also an interpolation method used where there are two or more spatially interdependent variables. Often, cokriging is used when one or more other properties have been extensively sampled in comparison to the variable of interest (Oliver, 1987). Ideally, the densely sampled variable, called a covariate, secondary variable or subsidiary variable, is measured more cheaply and quickly than the property of interest, or target variable.

Measuring soil variability through the use of EMI can be performed quickly and inexpensively. Soil electrical conductivity (EC) can be measured on an extremely small grid in a rapid, easy and nondestructive manner by the use of EMI. Soil EC is a measurement that is affected primarily by a combination of soil water content, dissolved salt content, clay content and mineralogy and soil temperature (McNeill, 1980a). Soil EC measured by EMI has been correlated with clay content (Williams and Hoey, 1987; Doolittle et al., 1994), soil water content (Kachanoski et al., 1988; Sheets and Hendrickx, 1995), sand deposition (Kitchen et al., 1996), total soluble salts (Williams and Hoey, 1987), yield (Jaynes et al.,

1995) and soil available N (Eigenberg et al., 2002). Therefore, soil EC data may serve as a covariate and noninvasively provide valuable and inexpensive information to aid in the production of more accurate soil maps for certain variables (Jaynes, 1996).

In this study, cokriging methods were compared to kriging methods for measured soil properties. The objective of this study was to determine if the use of soil EC as a covariate improved mapping accuracy of five soil variables across four sampling schemes and two sampling densities in a central Iowa pasture.

Materials and Methods

Research was conducted at the Iowa State University Rhodes Research Farm (41°52′N, 93°10′W) in central Iowa. The Wisconsin loess-covered landscape has an underlying Yarmouth-Sangamon paleosol. The soils are primarily slope and erosion phases of the Fayette and Clarinda series (Fenton, 2001). The pasture site of study included topographically distinct summit, sideslope, toeslope, backslope and opposite summit landscape positions (Fig. 1).

A dense sampling grid consisting of 116 points was devised for a 0.42 ha, nongrazed grass-legume pasture (Fig. 2). Sampling points were arranged in a triangular grid with inter- and intra-row separation distances of 6 m. In order to obtain data from samples located closer than 6 m, an additional point was sampled within each row at randomly chosen 1 or 2 m separation distances. This short range variation in soil samples was investigated in order to obtain a more reliable experimental semivariogram model (Burgess and Webster, 1980; Kravchenko and Bullock, 2002).

Apparent soil EC was measured with the Geonics® EM-38 as it was pulled behind a four-wheel drive vehicle in a nonconductive cart (Fig. 3). The EM-38 was operated in the vertical dipole orientation at 0.20 m above the soil surface. The EM-38 integrates over an area approximately equal to its length of 1 m and over a depth of approximately 3 m

(McNeill, 1980b). However, the measurement is primarily influenced by the 0 to 1.5 m depth increment (McNeill, 1980b). The vertical rather than horizontal dipole orientation was used so that any fluctuations in carrying height of the EM-38 above the soil surface would have little impact on EC readings (McNeill, 1980b). Given the speed of travel and rate of EC data collection, an EC reading was logged approximately every 4 m in 2 m transects throughout the pasture. Each soil EC measurement was georeferenced using a Trimble Global Positioning Systems (GPS) Pathfinder® Pro XR receiver, and all GPS locations were differentially corrected (DGPS) to obtain 1-2 m accuracy. EM-38 measurements were recorded via a direct connection to Trimble System Controller (TSC) 1 Asset Surveyor software in the GPS datalogger. Positional data for the soil EC values were corrected for the lag distance between GPS receiver and the EM-38 instrument. Soil EC values at each of the 116 grid points were interpolated from the dense data set of the 834 georeferenced EC points.

Soil samples were collected via coring following the EM-38 measurements. At each of the 116 sampling sites, five 15 cm soil cores were collected and combined for soil pH and available P and K analysis in the Iowa State University Soils Testing Laboratory (Ames, IA). Organic matter was analyzed from a single core composed of three depth increments to 51 cm, using a dry combustion method. Average organic matter percentage over all three depth increments was reported. To quantify soil moisture, a single core composed of seven 15 cm samples was taken. Each 15 cm sample was analyzed using the gravimetric moisture method (Buckman and Brady, 1971), and an average percent soil moisture was reported.

Elevation data were recorded using a Corvallis Microtechnology, Inc. (CMT) real time kinematic (RTK) system, and slope data were calculated from this using ArcView 3.2 Spatial Analyst (ESRI, 1996). Geostatistical analyses were performed using ArcView 8.1 ArcGIS Geostatistical Analyst (ESRI, 2001).

Sampling Schemes

Four different sampling patterns at two densities were created from the original sampling grid (n = 116). The sampling schemes were a grid pattern, a triangular pattern, a multistage clustering scheme and a random scheme. The n = 30 and n = 15 sampling schemes are shown in Figures 7 and 8, respectively. Because the sampling schemes were created from the original sampling grid, there were some restrictions on the arrangement of the patterns. The grid pattern was a rectangular grid with 6 m intra-row and 12 m inter-row separation distances for the n = 30 scheme. Intra-row grid spacing was also 6 m for the n = 15 scheme, but inter-row separation distance increased to 24 m. The sampling scheme originated on the west end of the pasture and because of the specified sample size, sampling density on the east end of the pasture was less dense for both the n = 15 and n = 30 grid schemes. For similar restriction reasons, the n = 30 triangular scheme is more dense on the east end of the pasture and triangular grid size varies for the n = 15 scheme. The triangular pattern was not equilateral; the triangles were formed with a base length between points of 6 m and a side length of 19.0 m for nearly all of the pasture and 12-13.4 m on the extreme east end of the pasture for the n = 30 scheme. Base length between points was also 6 m for the n = 15 triangular scheme with side lengths of 30.6 m and 42.4 m. The multistage clustering scheme was produced by first implementing the fuzzy k-means algorithm. The algorithm was used to stratify the field into relatively homogeneous zones based on densely collected soil parameters. The fuzzy k-means method has been utilized for classifying soil and landscape data when binary or strictly discrete groupings are not adequate to describe natural systems (Burrough, 1989; McBratney and de Gruijter, 1992; Odeh et al., 1992; Irvin et al., 1997). Given the continuous nature of soils, fuzzy set classification provides a suitable means of classifying areas of a field. A combination of topographic attributes and apparent soil EC were used to delineate zones using the software program Management Zone Analyst© (University of Missouri-Columbia and Agricultural Research Service, 2000). The

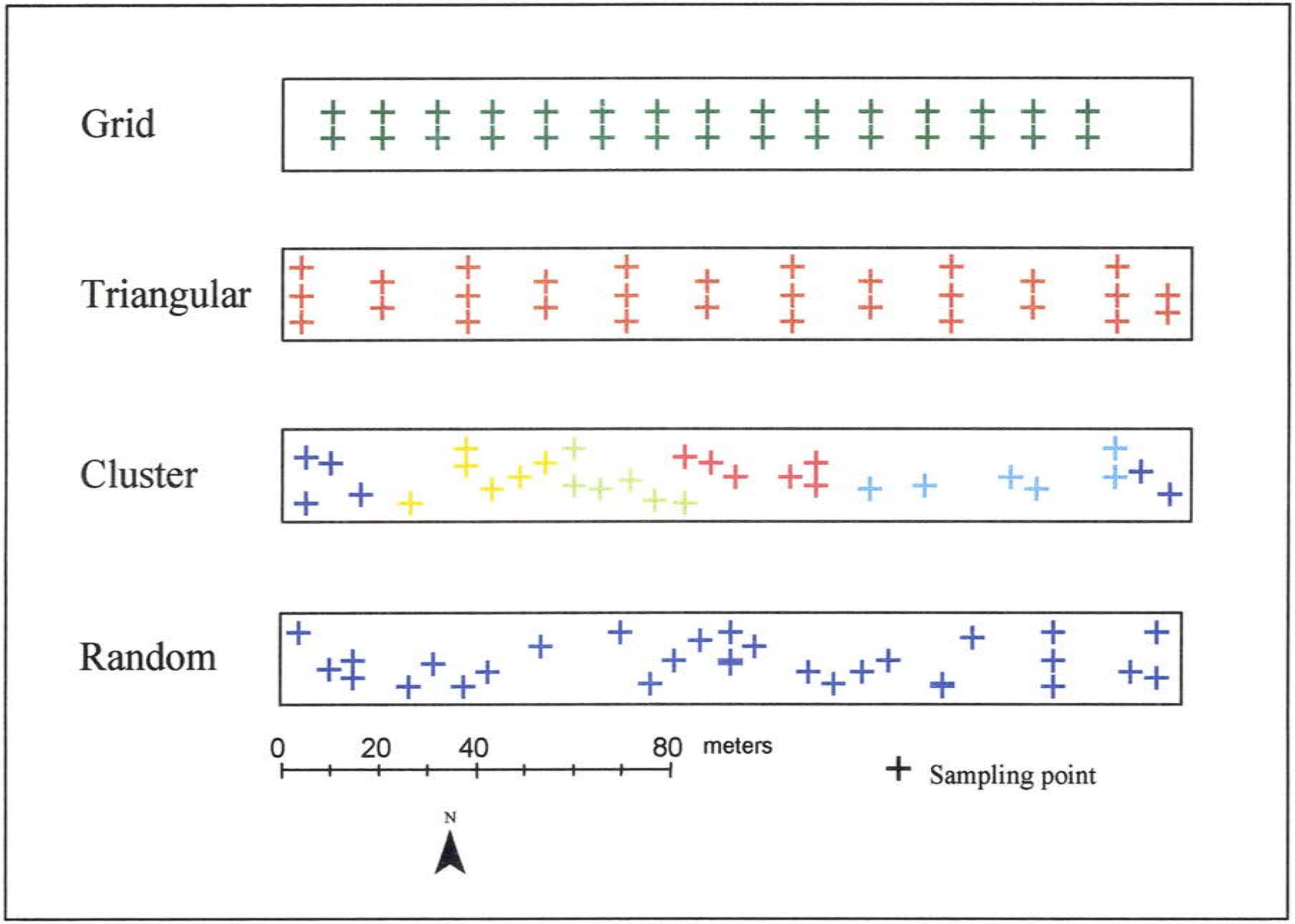

Figure 7. Four sampling schemes at n = 30 density.

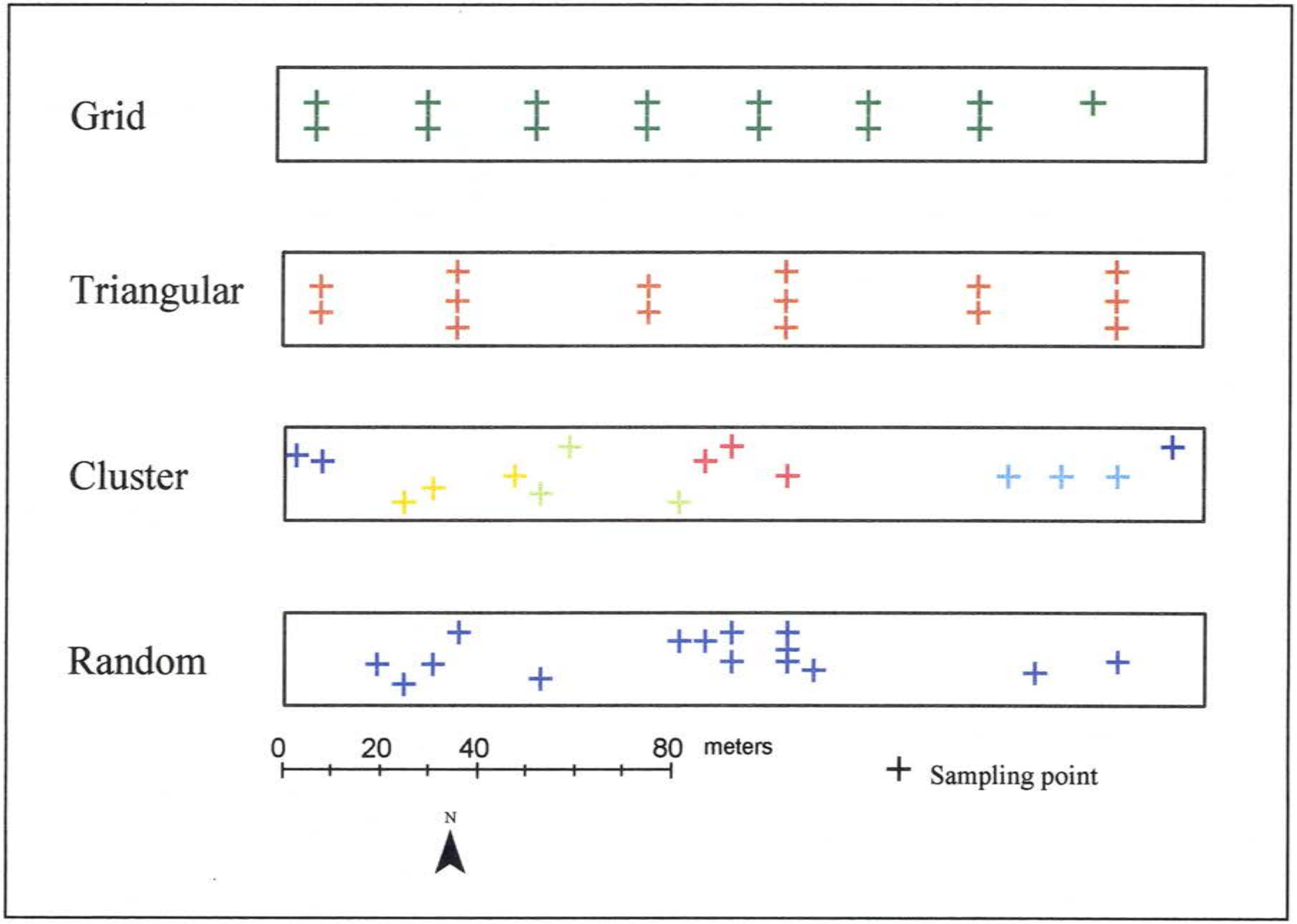

Figure 8. Four sampling schemes at n = 15 density.

clustering algorithm was iterated for 3-9 zones. Based on the fuzziness performance index (FPI) and normalized classification entropy (NCE), five zones appeared optimal for establishing strata of homogeneity in the pasture. The FPI and NCE performance indices were discussed and used to evaluate the varying number of climatic classes by McBratney and Moore (1985). Within these five strata, two intensities of a ranked set sampling scheme (McIntyre, 1952) were analyzed: n = 30 and n = 15 (Figs. 7 and 8). Ranked set sampling was first described by McIntyre (1952) as a method for obtaining more precise and unbiased measurements of forage yield. The five strata delineated by clustering corresponded to the sets in ranked set sampling. The points in each of the five strata were ranked in order of EC value magnitude. Soil EC was chosen as a concomitant variable because it is the variable most easily and accurately collected (Patil et al., 1994), and it was correlated with the soil variables of interest. When constructing the n = 30 scheme, six points were selected in each of the five strata. With this scheme, the six points were chosen based on maximizing the within-zone variation of soil EC. The six points were selected based on choosing the minimum, maximum, and four in-between quantile values of soil EC within each zone. Maximum within-zone variation was sought in order to maintain the variability identified throughout the field. Similarly, when devising the n = 15 scheme, three points were selected in each of the five strata with the goal of maximizing the within-zone variation of the concomitant variable, soil EC. The three points were selected based on choosing the minimum, maximum and median values of soil EC within each zone. Lastly, using a random number generator, sampling schemes with sizes n = 30 and n = 15 were produced from the original 116 sampling points for the random sampling schemes.

Because direct soil measurements were taken at each of the original 116 sampling points, relatively large validation sets were available. Eighty-six and 101 points, respectively, were used as independent validation sets for the n = 30 and n = 15 sampling densities.

Results and Discussion

Statistical Data Analysis

Statistical summaries of topographic and soil attributes for the initial, dense grid (n = 116) sampling scheme are presented in Table 1. With a range in elevation of 6.4 m, a CV of 0.7 did not describe the actual variability in topography. The CV was very small due to the relatively large mean value for elevation as compared to the SD. Large CVs and ranges for soil P, K, pH, OM and moisture illustrated the magnitude of soil variability within the pasture (Table 1). From a sampling perspective, this known variability is of interest because it is important that a sampling technique can identify this variability in its resulting map. Furthermore, management decisions are made based upon this map.

The degree to which soil EC is spatially correlated with the other soil variables of interest is called coregionalization. Although statistical correlation does not imply spatial correlation, the dense data set in this study does suggest a baseline for spatial relationships to exist among the soil variables. The relationships among the chemical and physical attributes for the initial, dense sampling scheme are shown in Table 2. The results of the large database indicated a strong correlation between soil EC and soil pH, elevation and soil K (r values greater than 0.5). Soil EC was moderately correlated with soil moisture and OM values ($0.10 < r < 0.50$) and weakly correlated with soil P and slope ($r \leq |0.10|$) (Table 2).

In the pasture of study, variability in soil EC appeared closely related to the variation in landscape position and depth to paleosol (Fenton, 2002). Higher values of soil EC were measured in the toeslope positions. These positions have a higher moisture content and are underlain by a clay-textured paleosol (Clarinda series). Lower values of soil EC were measured upslope on the summit positions where the soils were developed entirely in loess. Middle values were measured on the backslope where the soils are formed in loess and the underlying paleosol. Because the soil was not dominated by carbonates, the variation in soil EC values was likely related to soil moisture content and textural properties (Brevik and

Fenton, 2002). Textural properties influence soil parameters such as organic matter and ionic properties; thus, it was concluded that soil EC is measuring variation in soil properties related to moisture and texture.

Geostatistical Data Analysis

Ordinary kriging is widely used in soil literature as a method for interpolation. Both the theory and application of ordinary kriging are described in depth by Journel and Huijbregts (1978) and McBratney and Webster (1983a). We investigated the value of using soil EC as a covariate for improving mapping accuracy by comparing kriging the soil parameters to cokriging the soil parameters with soil EC as a covariate.

Kriging and cokriging were performed using the Geostatistical Analyst extension in ArcView 8.1 (ESRI, 2001). Adequacy of the chosen variogram models was tested using cross-validation (Vauclin et al., 1983; Warrick et al., 1986). In a cross-validation, each point in the sampling scheme is removed singly and its value is predicted based on kriging the remaining data. The resulting root mean square error (RMSE) of the cross-validation process was examined, and the variogram model with the lowest RMSE was selected (Vauclin et al., 1983; Heisel et al., 1999). Skewness results indicated that the data was not normally distributed. To improve normality, the data for soil organic matter, phosphorus and potassium was log transformed. Data were reported on the non-transformed values.

The pasture of study was oriented mostly in one dimension, and there were insufficient sample pairs of the soil variables for the n = 30 and n = 15 sampling schemes to obtain well-structured directional semivariograms (Trangmar et al., 1986). Therefore, it was assumed that all semivariograms were isotropic. Lag distances ranged from 2 m to 14 m with the majority of values being 8-9 m. Lag distances were autocalculated using the ArcView 8.1 Geostatistical Analyst extension (ESRI, 2001). This method tries a series of lag values, with their size increasing in a geometric sequence. Geostatistical Analyst then looks through all the lags and finds the lag and set of variogram parameters that have the "best fit", or smallest weighted least squares (Ver Hoef, 2002).

The five soil variables of interest, soil P, K, pH, OM and moisture were undersampled compared to the soil EC readings. This is the situation where cokriging is most useful. Because of the ease of collecting dense, rapid and georeferenced soil EC data, its spatial relationship with the five soil parameters was explored. In this 0.42 ha pasture, 834 soil EC points were measured. Consequently, the ratio of sampling intensities of soil EC to the other soil variables was nearly 28 for the n = 30 schemes and nearly 56 for the n = 15 schemes (Fig. 9).

To apply cokriging it was necessary to model semivariograms for each soil variable separately as well as cross-variograms for all pairs of soil EC and soil variable measured at the same location (McBratney and Webster, 1983a; McBratney and Webster, 1983b; Vauclin et al., 1983; Triantafilis et al., 2001). The correlations shown in Table 2 were also examined (Chien et al., 1997). Soil variables that were more highly correlated with soil EC also produced better-structured cross-variograms (Journel and Huijbregts, 1978). Proof of some level of coregionalization between soil EC and the five soil parameters of interest was important for further cokriging steps.

Map Results

Kriging vs. Cokriging

In general, a visual discrepancy existed between the kriged and cokriged maps. The cokriged maps exhibited more local detail and less smoothness in their depiction of the variability of the five soil parameters. However, both methods generated similar trends in the variability of the soil parameters. Comparable results have been found for topsoil silt cokriged with subsoil silt and sand as covariates (McBratney and Webster, 1983a), $NaHCO_3$-extractable P cokriged with 25% HCl-extractable P (Trangmar et al., 1986), NO_3 cokriged

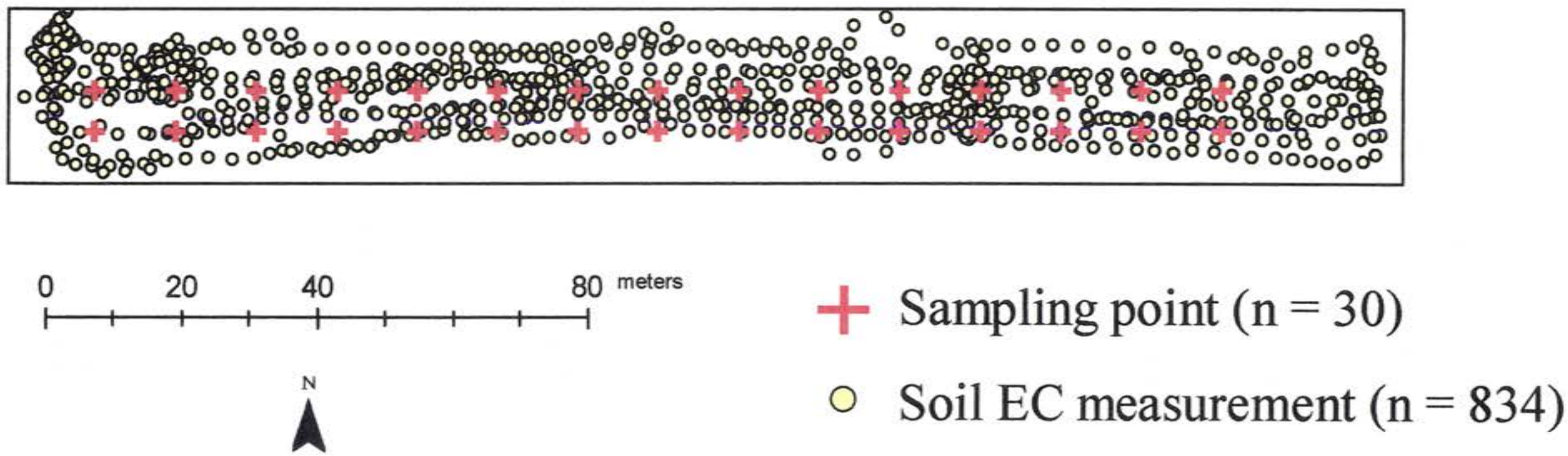

Figure 9. Map of soil sampling points and soil EC covariate points.

with soil EC (Zhang et al., 1992) and soil EC of soil paste extract cokriged with apparent soil EC (Vaughan et al., 1995). Improved local detail of the cokriging maps was due to the finer sampling grid of the covariate, soil EC (McBratney and Webster, 1983a). Representative maps of these observations are shown in Figure 10.

Several exceptions to cokriging's trend in increased local detail of variation were found in mapping of soil moisture, OM and P. However, these exceptions all exhibited similar spatial variability. In these anomalies, the autocorrelation of the soil variables revealed small lags and short range values; thus, the soil variables displayed short-range spatial variability. Small lags ranged from 2-4 m and range values were only 14-30 m for the semivariograms of the soil variables (data not shown). In addition, it is interesting to note that the three soil variables that exhibited less local detail with cokriging were also the soil variables least correlated with soil EC (Table 2). This result could be due to the relatively low cross-correlation between these variables (soil moisture, OM and P) and soil EC; thus, modeling of the cross-variogram was made difficult (McBratney and Webster, 1983a). Or, the result could be due to the geometric arrangement of the soil sampling points themselves and how that lended to modeling semivariograms and cross-variograms (McBratney and Webster, 1983a). Whether or not the reduced local detail in the maps actually affected mapping accuracy will be discussed in the quantitative portion of the results.

Sampling density comparison

In general, maps of n = 30 and n = 15 schemes displayed similar spatial patterns for both kriging and cokriging. Visually, it appeared that cokriging the soil variables of interest with soil EC had a greater impact on the maps for the less dense sampling scheme. The cokriged maps for n = 15 exhibited greater local detail than the cokriged maps for n = 30, especially for the soil variables most highly correlated with soil EC (soil pH and K) (Table 2). Thus, when there are fewer points measured invasively, the variability of the noninvasive

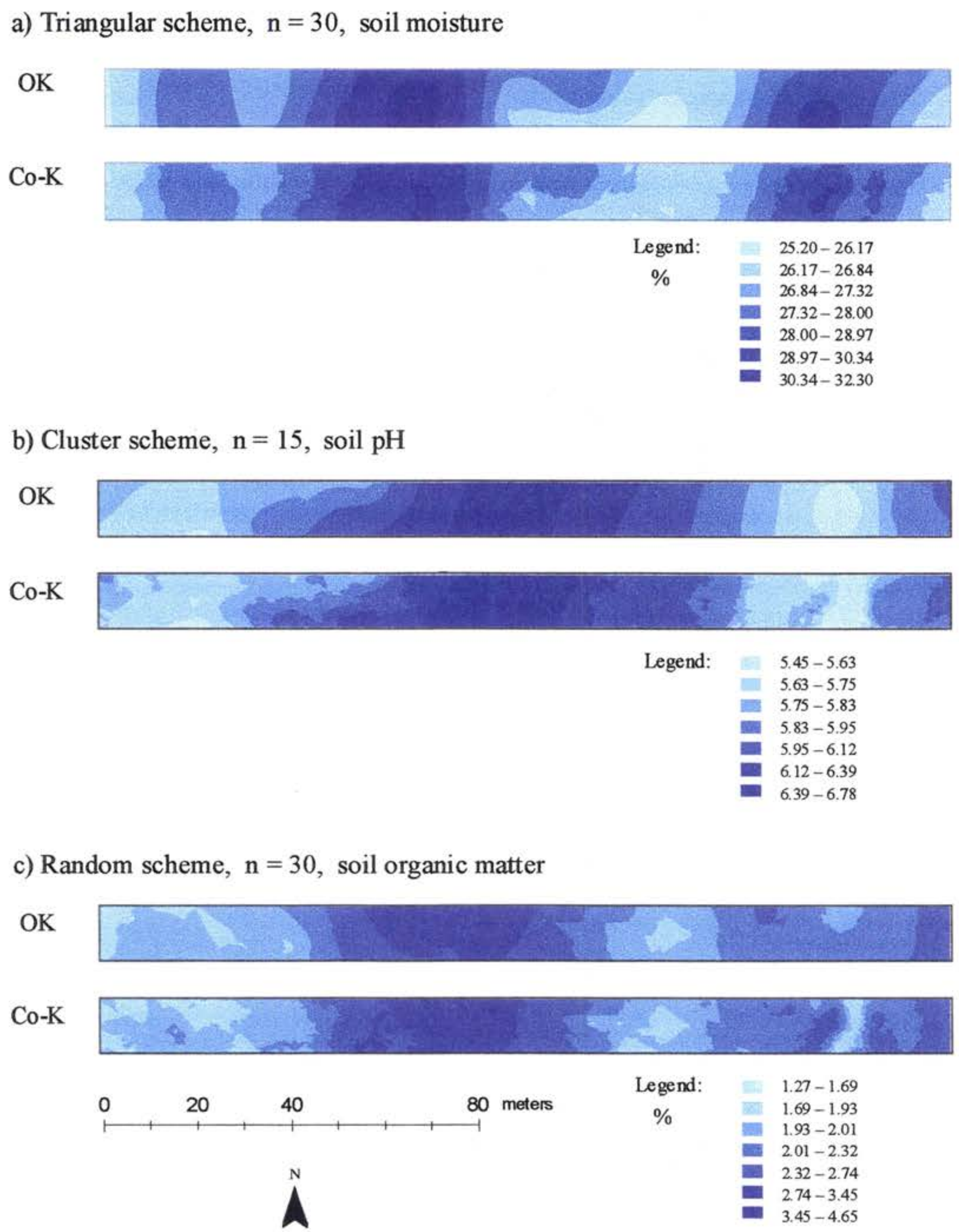

Figure 10. Representative comparisons of kriged (OK) and cokriged (Co-K) maps for a) n = 30 triangular sampling scheme, soil moisture, b) n = 15 cluster sampling scheme, soil pH and c) n = 30 random sampling scheme, soil organic matter.

soil parameter (soil EC) appeared to have a greater influence on prediction of unsampled areas of soil pH and K.

Few other observations could be made from visual assessment of the kriged and cokriged maps. Differences in mapping characteristics of the best-fit semivariogram models made it difficult to conclude “real” differences among maps. Maps are effective for identifying trends, but the actual differences between methods should be analyzed quantitatively.

Sampling scheme comparison

Based on the design and results of this experiment, it cannot be concluded which of the four sampling schemes is optimal for producing accurate maps of soil variability. Each of the schemes used in this study was just one depiction of a grid scheme or one depiction of a random or clustered scheme in the pasture. Multiple scenarios of each scheme should be analyzed further in order to answer this question. General conclusions about the sampling schemes will be discussed after examining them quantitatively.

Quantitative Results

Kriging vs. Cokriging

When comparing results estimated with kriging and cokriging, two parameters were analyzed. Root mean square error (RMSE) of prediction and the correlation between predicted values from kriging and the actual values taken from invasive soil measurements were evaluated. As previously mentioned, validation sets of 86 and 101 sample points were available for the $n = 30$ and $n = 15$ sampling schemes, respectively. The RMSE should be small for an unbiased and precise prediction. For both kriging and cokriging at each sampling density and for each sampling scheme, the root mean square error was calculated as:

$$RMSE = \left\{ \frac{1}{n} \sum_{i=1}^{n} [z(s_i) - z^*(s_i)]^2 \right\}^{0.5}$$

where n is the number of sample sites in the validation set, $z(s_i)$ are the observed soil values and $z^*(s_i)$ are the predicted values. As shown in Table 9, RMSE for cokriging was not consistently lower than the RMSE for kriging as would be expected if cokriging helped improve the prediction of the validation sites. Correlation between each soil variable and soil EC was likely not the single factor affecting the success of cokriging. However, in general, higher correlation between the target soil variable and soil EC resulted in a greater reduction in RMSE (Fig. 11). Relative reduction in RMSE was defined by:

$$100(RMSE_k - RMSE_{ck})/RMSE_k$$

where $RMSE_k$ and $RMSE_{ck}$ are the root mean square errors of kriging and cokriging, respectively (adapted from Zhang et al., 1992). These findings were comparable to the observations by Yates and Warrick (1987), which found that a reduction in kriging variance is observed as the correlation between the covariate and target variable increased. The findings of the current study were more volatile that those found by Yates and Warrick (1987), but Figure 11 shows that the lower correlation between soil EC and P resulted in the lowest reductions (largest gains) of RMSE. While OM exhibited the largest reduction in RMSE, soil pH showed more consistent reductions in RMSE for the sampling scheme and density combinations (Fig. 11). General trends can be drawn from this figure, but it is important to note that the efficacy of cokriging is more than a function of correlation between a covariate and target variable. The efficacy also includes the strength of the spatial cross-correlation between the covariate and target variable, the geometric sampling pattern and the ratio of the sampling intensities of the covariate and target variable (McBratney and Webster, 1983a).

Table 9. Validation set root mean square errors (RMSE) of soil data prediction for kriging and cokriging the n = 30 and n = 15 sampling densities for each of the four sampling schemes. Coefficient of determination (r^2) between predicted and actual values for validation set.

			Grid		Triangular		Cluster		Random	
Target variable	Sample size		OK†	Co-K	OK	Co-K	OK	Co-K	OK	Co-K
P	30	RMSE‡	5.468	5.516 (+)§	5.575	5.569 (-)	4.903	6.787 (+)	6.508	6.929 (+)
		r^2	0.470	0.262 (-)	0.332	0.340 (+)	0.549	0.14 (-)	0.206	0.088 (-)
	15	RMSE	8.106	8.071 (-)	7.693	7.625 (-)	6.884	7.382 (+)	8.008	7.952 (-)
		r^2	0.031	0.057 (+)	0.022	0.026 (+)	0.265	0.046 (-)	0.023	0.028 (+)
K	30	RMSE	52.92	55.09 (+)	43.08	39.85 (-)	45.7	44.13 (-)	45.88	45.71 (-)
		r^2	0.389	0.371 (-)	0.447	0.499 (+)	0.518	0.500 (-)	0.516	0.509 (-)
	15	RMSE	47.27	49.03 (+)	46.79	46.98 (+)	44.16	45.28 (+)	62.39	61.49 (-)
		r^2	0.411	0.357 (-)	0.491	0.479 (-)	0.458	0.437 (-)	0.099	0.111 (+)
pH	30	RMSE	0.178	0.176 (-)	0.185	0.179 (-)	0.169	0.165 (-)	0.175	0.168 (-)
		r^2	0.828	0.835 (+)	0.830	0.845 (+)	0.858	0.854 (-)	0.845	0.854 (+)
	15	RMSE	0.184	0.186 (+)	0.229	0.204 (-)	0.306	0.286 (-)	0.224	0.235 (+)
		r^2	0.817	0.806 (-)	0.795	0.802 (+)	0.555	0.62 (+)	0.795	0.731 (-)
OM	30	RMSE	0.822	0.693 (-)	0.746	0.744 (-)	0.762	0.769 (+)	0.767	0.763 (-)
		r^2	0.116	0.191 (+)	0.232	0.235 (+)	0.187	0.166 (-)	0.145	0.159 (+)
	15	RMSE	0.918	0.915 (-)	1.048	1.047 (-)	0.733	0.727 (-)	0.839	0.763 (-)
		r^2	0.016	0.020 (+)	0.036	0.037 (+)	0.250	0.257 (+)	0.012	0.129 (+)
Moisture	30	RMSE	2.017	1.957 (-)	2.004	1.991 (-)	1.611	1.539 (-)	2.536	2.559 (+)
		r^2	0.379	0.395 (+)	0.381	0.396 (+)	0.501	0.369 (-)	0.097	0.125 (+)
	15	RMSE	2.081	2.098 (+)	2.108	2.1 (-)	2.414	2.392 (-)	2.325	2.329 (+)
		r^2	0.208	0.201 (-)	0.213	0.214 (+)	0.044	0.080 (+)	0.131	0.109 (-)

† OK is ordinary kriging, Co-K is cokriging.

‡ RMSE units are specific to each soil variable: parts per million soil P and K, units pH, % OM and moisture.

§ Plus or minus sign in parentheses indicates an increase (+) or decrease (-) in RMSE or r^2 value between OK and Co-K.

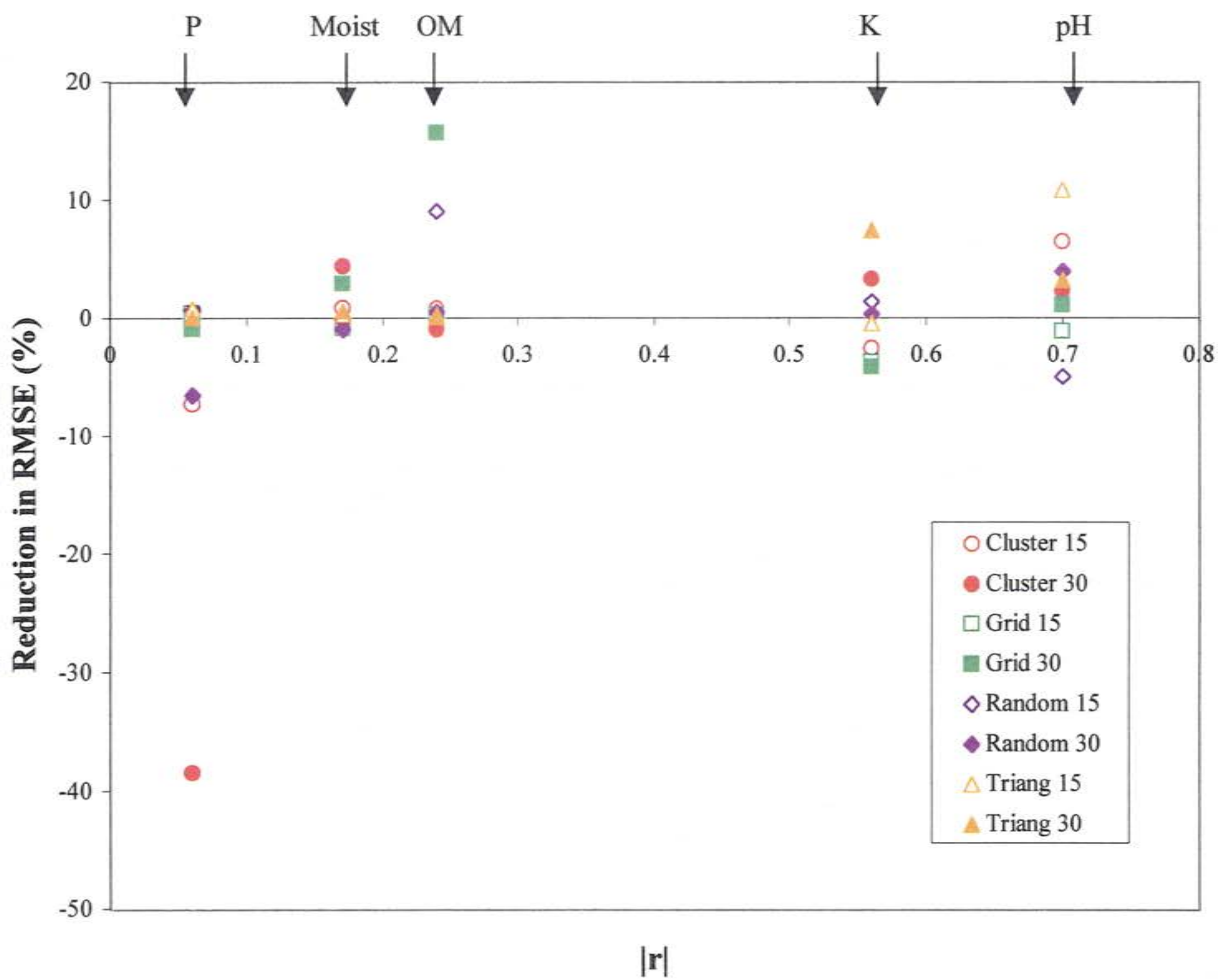

Figure 11. Reduction in root mean square error (RMSE) of prediction due to cokriging as a function of the absolute correlation (r) between soil EC and the five selected soil variables (P, moisture, OM, K and pH). Data is plotted for each soil sampling pattern and density.

In addition, r^2 values did not increase due to cokriging for every sampling scheme and density combination (Table 9). Soil OM, pH and moisture displayed improved r^2 values most consistently among the soil variables.

Sampling density comparison

By decreasing the number of sampling sites, Vauclin et al. (1983) demonstrated the advantage of cokriging over kriging by comparing the estimation variances at interpolated points. When the effect of sampling density was compared to cokriging efficacy, no clear relationship was found (Table 9). For example, cokriging with soil EC provided better prediction accuracy for soil P at the lower sampling density (n = 15) but better prediction accuracy for soil moisture at the higher sampling density (n = 30) (Fig. 11). It is hypothesized that the benefit of cokriging was not observed because of the random choice of fifteen points or because semivariogram and cross-variogram models were not well defined with such few sample pairs at a low density (n = 15). Absolute values of RMSE were generally higher and correlation between predicted and actual values were generally lower with the lower sampling density, as also shown by Zhang et al. (1992) (Table 9).

Sampling scheme comparison

The mapping accuracy from cokriging was also affected by the sampling pattern that was used. The triangular set of points showed the largest benefit from cokriging based on the reduction in RMSE (Fig. 11) and improvement in r^2 values (Table 9). The triangular scheme also showed the strongest response to correlation between soil EC and the soil variables. As shown in Figure 11, the reduction in RMSE increased favorably as the correlation between the covariate (soil EC) and the soil variables of interest increased. The cluster and random sampling schemes may have not have benefited as much from the cokriging with soil EC as a covariate because of the geometric pattern of the sampling sites. As seen from the maps of the cluster and random schemes (Figs. 7 and 8), coverage of the sampling is somewhat uneven. Because of these gaps in coverage, estimation of the unsampled sites may be difficult because information that could have been obtained is lacking (Webster and Oliver, 2001). The estimation error increases the further an interpolated point is from the sampling points (Burgess et al., 1981). In addition, because the soil variables were spatially autocorrelated to some degree, the closely neighboring points found in the random scheme were likely quite similar; thus, duplicated information was collected (Webster and Oliver, 2001). Despite gaps in sampling coverage, the benefit of soil EC as a covariate is that it is collected on a much denser grid than the soil variables of interest. However, because only one example of each sampling scheme was used, generalizations about each of the schemes cannot be made.

Conclusions

It is known that prediction accuracy of unsampled sites decreases as fewer points are sampled (Burgess et al., 1981). This reality is important from a field mapping and management perspective because it is impossible to sample every point in the field. Ironically, identifying field variability or homogeneity is central to precision agriculture and knowing what is going on in between the sampled points is just as important as knowing what is going on at the sampled points. Because soil EC is rapidly, easily and densely collected, it is possible that soil EC can help us to identify this in-between sampling point variability.

This study examined the use of soil EC as a covariate for improving map accuracy of five soil variables sampled at two densities from four different sampling schemes. Soil EC's spatial relationship with the five soil variables was exploited through the use of cokriging. Maps resulting from cokriging soil EC with each of the soil variables exhibited more local detail than the kriged maps of each soil variable. Whether this increased detail resulted in better mapping accuracy was an important concern.

It is difficult to establish a correlation value between the covariate and target variable that assures improved map accuracy with cokriging. Previous studies have suggested an absolute sample correlation greater than 0.5 (Yates and Warrick, 1987) while others have seen lower errors of estimation for cokriging with a relatively low correlation coefficient of -0.38 (Leenaers et al., 1990). Varying ratios of covariate to target variable are also possible and modeling the cross-correlation between covariate and target variable is important. Thus, there are several requirements for a successful outcome from cokriging.

This study showed an overall, but inconsistent, improvement in kriging variance and prediction accuracy of unsampled sites when cokriging was implemented. The improvement was generally greater for soil variables more highly correlated with soil EC. A clear improvement from cokriging with a lower density target variable was not observed. The triangular grid appeared to benefit most from cokriging, but the results are inconclusive because only one triangular scheme (not multiple) was analyzed.

Soil EC is a rapidly, densely, noninvasively and easily collected method for quantifying soil variability. Integrating this broad base of information with less densely and invasively collected soil parameters can help us to better characterize pasture soil variability. Consequently, this could lead to more accurate maps and more precise management of field inputs such as nutrients and lime.

References

Brevik, E.C., and T.E. Fenton. 2002. Influence of soil water content, clay, temperature, and carbonate minerals on electrical conductivity readings taken with an EM-38. Soil Surv. Horiz. 43:9-13.

Buckman, H.O., and N.C. Brady. 1971. The nature and properties of soils. 7th ed. Macmillan, New York.

Burgess, T.M., and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil Sci. 31:315-331.

Burgess, T.M., R. Webster, and A.B. McBratney. 1981. Optimal interpolation and isarithmic mapping of soil properties. IV. Sampling strategy. Journal of Soil Science 32:643-659.

Burrough, P.A. 1989. Fuzzy mathematical methods for soil survey and land evaluation. Journal of Soil Science 40:477-492.

Chien, Y., D. Lee, H. Guo, and K. Houng. 1997. Geostatistical analysis of soil properties of mid-west Taiwan soils. Soil Science 162:291-298.

Doolittle, J.A., K.A. Sudduth, N.R. Kitchen, and S.J. Indorante. 1994. Estimating depths to claypans using electromagnetic induction methods. J. Soil and Water Cons. 49:572-575.

Eigenberg, R.A., J.W. Doran, J.A. Nienaber, R.B. Ferguson, and B.L. Woodbury. 2002. Electrical conductivity monitoring of soil condition and available N with animal manure and a cover crop. Agriculture, Ecosystems and Environment 88:183-193.

ESRI. 1996. ArcView Spatial Analyst, Redlands, CA.

ESRI. 2001. ArcView 8.1. Release ArcGIS Geostatistical Analyst, Redlands, CA.

Fenton, T.E. 2001. Personal communication. Department of Agronomy, Iowa State University, Ames, Iowa.

Fenton, T.E. 2002. Personal communication. Department of Agronomy, Iowa State University, Ames, Iowa.

Golden Software, Inc. 1999. Surfer Version 7.0, Golden, Colorado.

Heisel, T., A.K. Ersboll, and C. Andreasen. 1999. Weed mapping with co-kriging using soil properties. Precision Agriculture 1:39-52.

Irvin, B.J., S.J. Ventura, and B.K. Slater. 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin. Geoderma 77:137-154.

Jaynes, D.B. 1996. Mapping the areal distribution of soil parameters with geophysical techniques. Application of GIS to the Modeling of Non-Point Source Pollutants in the Vadose Zone, SSSA Special Publication 48. SSSA, Madison, WI.

Jaynes, D.B., T.S. Colvin, and J. Ambuel. 1995. Yield mapping by electromagnetic induction. in Site -Specific Management for Agricultural Systems. ASA-CSSA-SSSA, Madison, WI.

Jensen, J.R. 1996. Introductory digital image processing: a remote sensing perspective. 2nd ed. Prentice Hall, Upper Saddle River, N.J.

Journel, A.G., and C.J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York.

Kachanoski, R.G., E.G. Gregorich, and I.J. Van Wesenbeeck. 1988. Estimating spatial variations of soil water content using noncontacting electromagnetic inductive methods. Can. J. Soil Sci. 68:715-722.

Kitchen, N.R., K.A. Sudduth, and S.T. Drummond. 1996. Mapping of sand deposition from 1993 midwest floods with electromagnetic induction measurements. J. Soil and Water Cons. 51:336-340.

Kravchenko, A.N., and D.G. Bullock. 2002. Spatial variability of soybean quality data as a function of field topography: I. Spatial Data Analysis. Crop Science 42:804-815.

Leenaers, H., J.P. Okx, and P.A. Burrough. 1990. Employing elevation data for efficient mapping of soil pollution on floodplains. Soil Use and Management 6:105-114.

McBratney, A.B., and R. Webster. 1983a. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. Journal of Soil Science 34:137-162.

McBratney, A.B., and R. Webster. 1983b. How many observations are needed for regional estimation of soil properties? Soil Science 135:177-183.

McBratney, A.B., and A.W. Moore. 1985. Application of fuzzy sets to climatic classification. Agricultural and Forest Meteorology 35:165-185.

McBratney, A.B., and J.J. de Gruijter. 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. Journal of Soil Science 43:159-175.

McIntyre, G.A. 1952. A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research 3:385-390.

McNeill, J.D. 1980a. Electrical conductivity of soils and rocks. Technical Note TN-5. Geonics, Ltd., Mississauga, Ontario, Canada.

McNeill, J.D. 1980b. Electrical conductivity of soils and rocks. Technical Note TN-6. Geonics, Ltd., Mississauga, Ontario, Canada.

Odeh, I.O.A., A.B. McBratney, and D.J. Chittleborough. 1992. Soil pattern recognition with fuzzy-c-means: Application to classification and soil-landform interrelationships. Soil Sci. Soc. Am. J. 56:505-516.

Oliver, M.A. 1987. Geostatistics and its application to soil science. Soil Use and Management 3:8-20.

Patil, G.P., A.K. Sinha, and C. Taillie. 1994. Ranked set sampling, p. 167-200, *In* G. P. Patil and C. R. Rao, eds. Environmental Statistics, Handbook of Statistics, Vol. 12. Elsevier Science, Amsterdam.

Sheets, K.R., and J.M.H. Hendrickx. 1995. Noninvasive soil water content measurement using electromagnetic induction. Water Resources Research 31:2401-2409.

Trangmar, B.B., R.S. Yost, and G. Uehara. 1986. Spatial dependence and interpolation of soil properties in West Sumatra, Indonesia: II. Co-regionalization and co-kriging. Soil Sci. Soc. Am. J. 50:1396-1400.

Triantafilis, J., I.O.A. Odeh, and A.B. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci. Soc. Am. J. 65:869-878.

University of Missouri-Columbia and Agricultural Research Service. 2000. Management Zone Analyst software. Release 1.0.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175-184.

Vaughan, P.J., S.M. Lesch, D.L. Corwin, and D.G. Cone. 1995. Water content effect on soil salinity prediction: a geostatistical study using cokriging. Soil Sci. Soc. Am. J. 59:1146-1156.

Ver Hoef, J.M. 2002. Personal communication. Alaska Department of Fish and Game, Fairbanks, Alaska.

Warrick, A.W., D.E. Myers, and D.R. Nielsen. 1986. Geostatistical methods applied to soil science, *In* A. Klute, ed. Methods of soil analysis, Part 1. 2nd ed. Agronomy 9:53-82.

Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists. John Wiley & Sons, Ltd., Chichester.

Williams, B.G., and D. Hoey. 1987. The use of electromagnetic induction to detect the spatial variability of the salt and clay contents of soils. Australian Journal of Soil Research 25:21-27.

Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. Soil Sci. Soc. Am. J. 51.

Zhang, R., D.E. Myers, and A.W. Warrick. 1992. Estimation of the spatial distribution of soil chemicals using pseudo-cross-variograms. Soil Sci. Soc. Am. J. 56:1444-1452.

CHAPTER 5. SPECTRAL REFLECTANCE AS A COVARIATE FOR ESTIMATING PASTURE PRODUCTIVITY AND COMPOSITION

Introduction

Sampling is the researcher's best way to learn about a population. When a pasture is considered to be a population, the task is to determine where to sample and how to assess the true variability as accurately as possible. With only a limited number of observations attainable due to time and labor constraints, interpolation is necessary to estimate values at unsampled points.

Matheron's theory of "regionalized variables" (1971) reported that field variables can be spatially correlated, or coregionalized. Geostatistics is the field of study that models spatial variability and is used to predict unknown values in space (Journel and Huijbregts, 1978). Consequently, spatial correlation within pastures is an opportunistic reality. Webster et al. (1989) capitalized upon this observation by designing a sampling scheme for ground-based radiometry measurements in both species-poor and species-rich grassland and winter barley. By fitting a semivariogram to radiation reflectance data sets, the error associated with estimating unsampled points for varying sampling intervals and densities was calculated (Webster et al., 1989).

The reflectance data measured in Webster et al.'s study (1989) was optimal in the sense that it was a spatially correlated variable that was rapidly and densely collected. In this study, the use of multispectral canopy reflectance data was used for similar reasons.

Multispectral reflectance measured using hand-held radiometers has been used to estimate many plant parameters of interest. Reflectance has been correlated with plant greenness in peanut (Nutter, 1989; Aquino et al., 1992) and in maize (Ma et al., 1996). Reflectance measurements were also found to be successful estimators of biomass in alfalfa (Mitchell et al., 1990), peanut (Nutter and Littrell, 1996) and potato (Bouman et al., 1992).

Seasonal biomass changes in tallgrass prairies were modeled using the normalized difference vegetation index (NDVI) along with several other environmental variables (Olson and Cochran, 1998). Light reflectance prior to anthesis may be able to predict grain yield in corn (Ma et al., 1996), and a good correlation was found between NDVI and millet total dry matter at harvest (Lawrence et al., 2000).

Plant disease progression has also been measured with multispectral radiometers. Disease severity in peanut canopies (Nutter, 1989), evaluation of fungicide efficacy in peanuts (Nutter et al., 1990) and occurrence of barley stripe disease (Nilsson and Johnsson, 1996) have been highly correlated with spectral reflectance using an eight-band hand-held radiometer. Because greenness is an indication of plant health and adequate fertilization, spectral reflectance has been incorporated to assess nitrogen fertilization of crops such as maize (Walburg et al., 1982; Ma et al., 1996; Blackmer et al., 1996).

Furthermore, the use of wavelength ratios can be used to discriminate between weed and crop species (Vrindts et al., 2002). Discriminant analysis in this study also resulted in 94 percent correct classification of broadleaved plants in test datasets of broadleaved plants and grasses (Vrindts et al., 2002).

Measuring pasture variability through the use of a ground-based multispectral radiometer can be performed quickly, nondestructively and inexpensively. Consequently, canopy reflectance data on a dense grid can be easily obtained. This dense data collection can be capitalized upon through the use of geostatistics. Kriging is a method of interpolation used when a variable displays spatial autocorrelation. Because reflectance values are spatially correlated (Webster et al., 1989), kriging can be used to predict reflectance at unsampled points. Cokriging is also an interpolation method used where there are two or more spatially interdependent variables. Often, cokriging is used when one or more other properties have been extensively sampled in comparison to the variable of interest (Oliver, 1987). Ideally, the densely sampled variable, termed a covariate, secondary variable or subsidiary variable, is measured more cheaply and quickly than the property of interest, or target variable. Therefore, canopy reflectance may serve as a covariate and noninvasively provide valuable and inexpensive information as a surrogate for prediction of other plant parameters of interest.

In this study, cokriging methods were compared to kriging methods for predicting measured plant parameters of interest. The objectives of this study were 1) to determine the relationships between easily collected canopy reflectance data and pasture biomass and species composition, and 2) to determine if the use of pasture reflectance data as a covariate improved mapping accuracy of biomass, percent grass cover and percent legume cover across three sampling schemes in a central Iowa pasture.

Materials and Methods

Research was conducted at the Iowa State University Rhodes Research Farm (41°52′N, 93°10′W) in central Iowa. The field of study was a 0.42 ha nongrazed, grass-legume pasture. The dominant species included smooth bromegrass (*Bromus inermis* L.), reed canarygrass (*Phalaris arundinacea* L.), Kentucky bluegrass (*Poa pratensis* L.), birdsfoot trefoil (*Lotus corniculatus* L.) and cicer milkvetch (*Astragalus cicer* L.). While there was mechanical removal of vegetation in early spring each year, the cool season grass-legume pasture had not been grazed in six years during a rotational grazing study. As an indication of field topography, the pasture site included topographically distinct summit, sideslope, toeslope, backslope and opposite summit landscape positions (Fig. 1). Slope ranged from approximately one to nine degrees and elevation ranged from 296.6 to 303.0 m above mean sea level.

A dense sampling grid consisting of 116 points was devised for a 0.42 ha nongrazed, grass-legume pasture (Fig. 2). Sampling points were arranged in a triangular grid with inter- and intra-row separation distances of 6 m. In order to obtain data from samples located closer than 6 m, an additional point was sampled within each row at randomly chosen 1 or 2 m separation distances. This short range variation in soil samples was investigated in order to obtain a more reliable experimental semivariogram model (Burgess and Webster, 1980; Kravchenko and Bullock, 2002). Each sampling point was georeferenced using GPS.

Canopy reflectance was measured at each of the 116 points in early June 2001. Grasses and legumes were in late vegetative/early reproductive stages. Canopy reflectance was measured using a CROPSCAN, Inc. (Rochester, MN) multispectral radiometer (model MSR87) over the center of each 1 m^2 plot (Fig. 12). Reflectance was measured at eight wavebands centered at: 460 nm, 510 nm, 560 nm, 610 nm, 660 nm, 710 nm, 760 nm and a wide infrared (IR) band from 1550-1750 nm. Bandwidths were approximately 10 nm for wavebands in the visible light and near infrared (NIR) regions (460-760 nm wavebands).

The circular field-of-view for the radiometer was matched as closely to the size of the 1 m^2 quadrat as possible. Because the diameter of the field-of-view was equal to ½ the height of the radiometer above the plant canopy, an average pasture canopy was measured. Two radiometer measurements were made in rapid succession and averaged at each of the 116 points. Measurements were taken between 1130 and 1400 h to minimize the effects of sun angle on incident radiation (Guan and Nutter, 2001). A bubble spirit level mounted on the support pole of the radiometer insured an appropriate angle (90°) of the radiometer sensors with respect to the pasture canopy. There were few clouds, a sunny sky and minimal wind on the date of measurement.

Following the reflectance measurements, species composition was evaluated. Using the Daubenmire canopy coverage method (1959), species within a 1 m^2 quadrat were ranked according to coverage abundance. Immediately following the ranking, the 1 m^2 quadrats were harvested using a mechanical hedge trimmer and cut as close to ground level as possible. Above ground biomass was placed in forced air drying ovens at 60°C degrees for 48 h to determine biomass for each plot.

Fig. 12. CROPSCAN, Inc. MSR87 multispectral radiometer in use.

Elevation data were recorded using a Corvallis Microtechnology, Inc. (CMT) real time kinematic (RTK) system, and slope data were calculated from this using ArcView 3.2 Spatial Analyst (ESRI, 1996). Geostatistical analyses were performed using ArcView 8.1 ArcGIS Geostatistical Analyst (ESRI, 2001).

Sampling Schemes

Three different sampling patterns of n = 30 were created from the original dense sampling grid (n = 116). The sampling schemes were a grid pattern, a triangular pattern and a random scheme. The sampling schemes are shown in Figure 13. Because the sampling schemes were created from the original sampling grid, there were some restrictions on the arrangement of the patterns. The grid pattern was a rectangular grid with 6 m intra-row and 12 m inter-row separation distances. The sampling scheme originated on the west end of the pasture and because of the specified sample size, sampling density on the east end of the pasture was less dense. For similar restriction reasons, the triangular scheme is more dense on the east end of the pasture. The triangular pattern was not equilateral; the triangles were formed with a base length between points of 6 m and a side length of 19.0 m for nearly all of the pasture and 12-13.4 m on the extreme east end of the pasture. Lastly, using a random number generator, a sampling scheme with size n = 30 was produced from the original 116 sampling points for the random sampling scheme.

Because plant measurements were taken at each of the original 116 sampling points, a relatively large validation set was available. Eighty-six points were used as an independent validation set for the n = 30 sampling density.

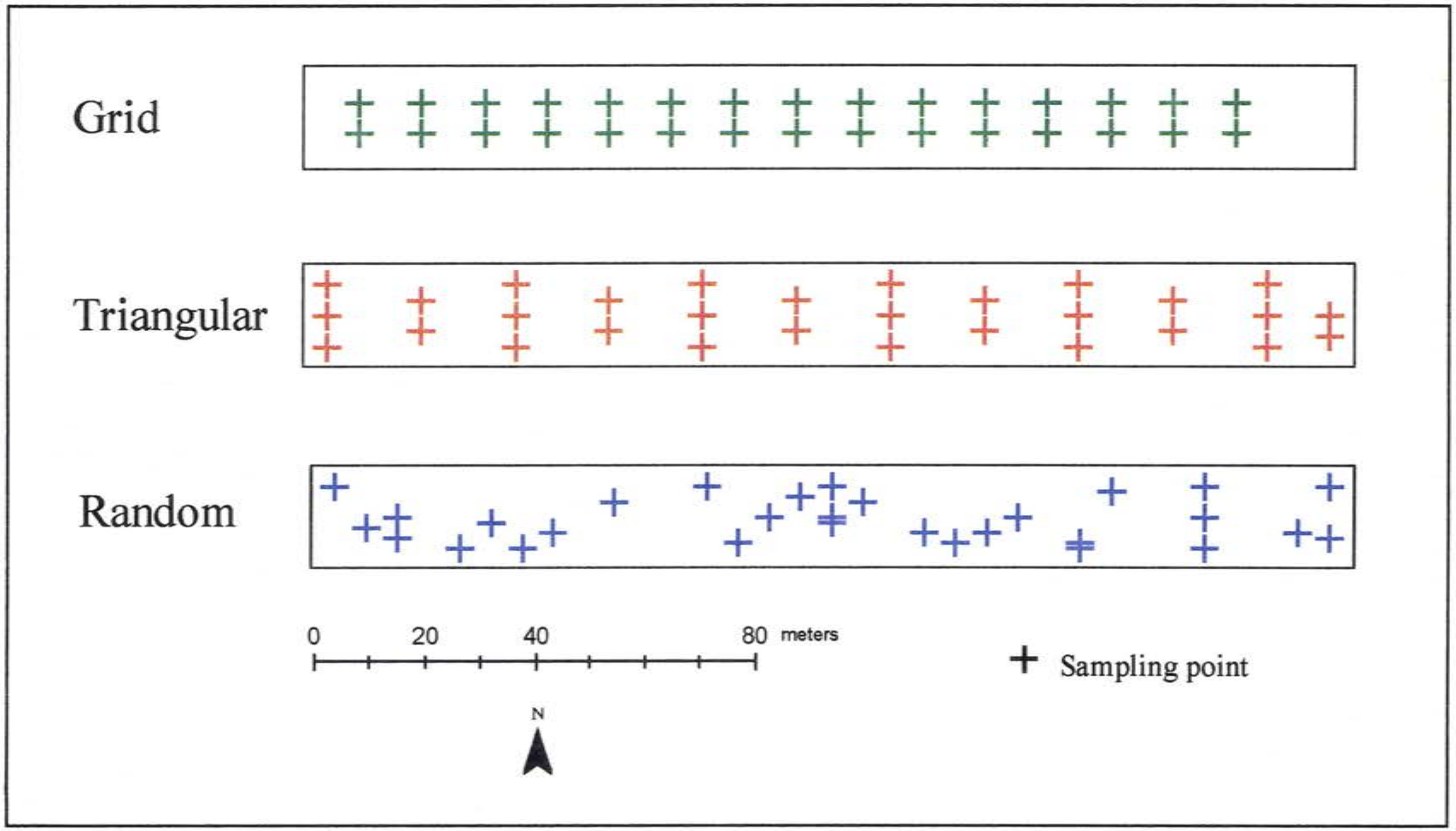

Figure 13. Three sampling schemes at n = 30 density.

Results and Discussion

Statistical Data Analysis

Summary statistics of pasture biomass and species composition values for the initial, dense grid (n = 116) sampling scheme are presented in Table 10. Biomass had a large range in values; however, it had a relatively low CV. On average, the pasture was composed of more grass than legume species, with smooth bromegrass as the most abundant grass. Absence of each species was found at one or more of the plots, as indicated by the minimum percent cover. However, grass was always present within the plots. The large CVs for species composition indicated the high variability in species occurrence throughout the pasture. From a sampling perspective, this known variability is of interest because it is important that a sampling technique can identify this variability in its resulting map. Furthermore, management decisions are made based upon this map. In addition, weed species were not particularly prevalent. The primary weed species were yellow nutsedge (*Cyperus esculentus* L.), common dandelion (*Taraxacum officinale* Weber) and yellow rocket (*Barbarea vulgaris* R. Br.).

The degree to which canopy reflectance values are spatially correlated with the plant parameters of interest is called coregionalization. Although statistical correlation does not imply spatial correlation, the dense data set in this study does suggest a baseline for spatial relationships to exist among the plant parameters. The relationships between reflectance values and productivity and species composition are shown in Table 11. Because both the 610 and 660 nm wavebands are in the red light region, ratios and indices were made with each of them where applicable. Red1 referred to 610 nm and Red2 was 660 nm. The normalized difference vegetation index was defined as:

$$\mathrm{NDVI} = \frac{NIR - red}{NIR + red}.$$

Table 10. Descriptive statistics of measured plant parameters for 116 sampling points.†

	Mean	SD	CV	Min.	Max.
Biomass, g‡	463	91	20	226	704
Grass, %	84.7	16.1	19.0	33.9	100.0
Smooth bromegrass, %	48.8	30.6	62.6	0	97.5
Kentucky bluegrass, %	7.3	8.6	118.4	0	39.5
Reed canarygrass, %	28.6	39.2	137.1	0	100.0
Legume, %	13.8	16.1	116.5	0	64.5
Birdsfoot trefoil, %	11.0	13.1	119.0	0	54.8
Cicer milkvetch, %	2.8	8.7	310.4	0	51.5
Other, %§	1.5	4.0	264.2	0	27.7

† Samples measured on 1 m^2 plots.

‡ Biomass is above ground plant material, % plant cover assessed by the Daubenmire (1959) canopy coverage method.

§ Indicates weed species.

Table 11. Partial correlation matrix among plant and canopy reflectance parameters for 116 sampling points.

	Grass	Leg.	Bio.	460	510	560	610	660	710	760	Wide	NDVI 1	NDVI 2	NIR/ Red1	NIR/ Red2	FIR/ Red1	FIR/ Red2
Grass, %†	1																
Legume, %	-0.97	1															
Biomass, g	-0.27	0.33	1														
460 nm	0.25	-0.28	-0.16	1													
510 nm	0.19	-0.22	-0.21	0.96	1												
560 nm	-0.08	0.06	-0.11	0.61	0.73	1											
610 nm	0.20	-0.23	-0.38	0.80	0.91	0.75	1										
660 nm	0.29	-0.31	-0.40	0.74	0.81	0.38	0.88	1									
710 nm	0.05	-0.09	-0.16	0.65	0.74	0.94	0.79	0.46	1								
760 nm	-0.07	0.07	0.30	0.08	0.005	0.40	-0.17	-0.51	0.39	1							
Wide IR	0.30	-0.27	-0.32	-0.11	-0.04	0.23	0.11	-0.05	0.24	0.20	1						
NDVI1	-0.19	0.19	0.43	-0.37	-0.47	-0.05	-0.64	-0.86	-0.08	0.86	0.07	1					
NDVI2	-0.23	0.23	0.40	-0.39	-0.46	0.04	-0.59	-0.87	-0.02	0.86	0.11	0.99	1				
NIR/Red1	-0.14	0.15	0.43	-0.35	-0.47	-0.10	-0.65	-0.84	-0.12	0.85	0.10	0.98	0.95	1			
NIR/Red2	-0.19	0.20	0.40	-0.40	-0.48	-0.03	-0.62	-0.86	-0.07	0.85	0.16	0.98	0.97	0.99	1		
FIR/Red1	0.11	-0.06	0.01	-0.65	-0.67	-0.35	-0.61	-0.66	-0.37	0.28	0.71	0.51	0.51	0.55	0.58	1	
FIR/Red2	-0.03	0.07	0.11	-0.65	-0.66	-0.19	-0.62	-0.79	-0.24	0.50	0.63	0.69	0.72	0.72	0.77	0.95	1

† Grass and legume measured by canopy coverage method (Daubenmire, 1959), biomass is above ground plant material, nm is nanometer, wide IR band is 1550-1750 nm, NDVI is normalized difference vegetation index calculated with Red1 (610 nm) or Red2 (660 nm), NIR is near infrared (760 nm), FIR is far infrared (1550-1750 nm).

The results of the large data set indicated several significant relationships between measured plant parameters of interest and reflectance values. For example, r values $\geq$ 0.40 were found between biomass and NDVI1, NDVI2, NIR/Red1 ratio and NIR/Red2 ratio (Table 11). Percent coverage of grass was correlated with reflectance in the wide IR band and at 660 nm (r values $\geq$ 0.29). Also, percent coverage of legume was correlated most highly with reflectance at 660 nm, 460 nm and at the wide IR band (r values $\geq |-0.27|$). These relationships were capitalized upon by cokriging. The spectral wavebands most highly correlated with the plant parameters of interest were used as covariates.

It is interesting to note that the correlation between percent grass cover and percent legume cover was -0.97. This highly significant value indicated that the occurrence of the two vegetation classes was inversely related. In addition, because of the fairly narrow spectrum of wavebands used and relatively wide spectral resolution (i.e. bandwidth), the reflectance values in the visible light region (400–700 nm) exhibited strong colinearity as did the ratios using the two wavebands of red light (Table 11).

Geostatistical Data Analysis

Interpolation is necessary to map a variable of interest at the ground from a sample of that variable. Kriging does this optimally in the sense that it estimates unsampled values with minimum variance. Both the theory and application of kriging are described in depth by Journel and Huijbregts (1978) and McBratney and Webster (1983a). We investigated the value of using one or more reflectance values or indices as a covariate for cokriging. Mapping accuracy of kriging the plant parameters of interest was compared to that of cokriging the plant parameters with reflectance values as a covariate.

Kriging and cokriging were performed using the Geostatistical Analyst extension in ArcView 8.1 (ESRI, 2001). Adequacy of the chosen variogram models was tested using cross-validation (Vauclin et al., 1983; Warrick et al., 1986). In a cross-validation, each point in the sampling scheme is removed singly and its value is predicted based on kriging the remaining data. The resulting root mean square error (RMSE) of the cross-validation process was examined, and the variogram model with the lowest RMSE was selected (Vauclin et al., 1983; Heisel et al., 1999). Skewness results indicated that not all the data were normally distributed. To improve normality, the reflectance data for the following covariates was log transformed: 660 nm, the wide IR band and the NIR/Red1 ratio. In addition, the $\sin^{-1}$ transformation was implemented for the NDVI1 and NDVI2 indices. Data were reported on the non-transformed values.

The pasture of study was oriented mostly in one dimension, and there were insufficient sample pairs of the plant parameters for the n = 30 sampling scheme to obtain well-structured directional semivariograms (Trangmar et al., 1986). Therefore, it was assumed that all semivariograms were isotropic. Lag distances ranged from 3 m to 16 m with the majority of values being 8 m. Lag distances were autocalculated using the ArcView 8.1 Geostatistical Analyst extension (ESRI, 2001). This method tries a series of lag values, with their size increasing in a geometric sequence. Geostatistical Analyst then looks through all the lags and finds the lag and set of variogram parameters that have the "best fit", or smallest weighted least squares (Ver Hoef, 2002).

The three plant parameters of interest, above ground biomass, percent grass cover and percent legume cover, were undersampled compared to the canopy reflectance readings. This is the situation where cokriging is most useful. Because of the ease of collecting dense, rapid and georeferenced canopy reflectance data, its spatial relationship with the three plant parameters was explored. In this 0.42 ha pasture, canopy reflectance at 116, 1 m^2 plots was measured. Consequently, the ratio of sampling intensities of reflectance to the other plant parameters was nearly 4:1 for the n = 30 scheme (Fig. 14).

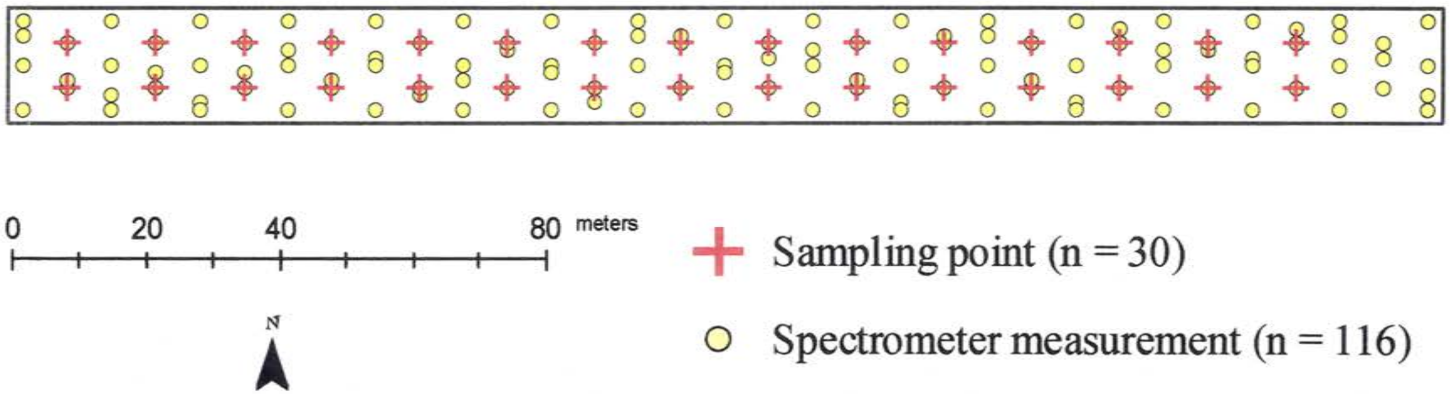

Figure 14. Map of plant sampling points and spectrometer covariate points.

To apply cokriging it was necessary to model semivariograms for each plant variable separately as well as cross-variograms for all pairs of canopy reflectance and plant parameters measured at the same location (McBratney and Webster, 1983a; McBratney and Webster, 1983b; Vauclin et al., 1983; Triantafilis et al., 2001). The correlations shown in Table 11 were also examined (Chien et al., 1997). Proof of some level of coregionalization between canopy reflectance and the three plant parameters of interest was important for further cokriging steps.

One, two and three covariates were examined to determine if using more than one covariate improved cokriging prediction accuracy (McBratney and Webster, 1983a). One covariate, $\sin^{-1}$ transformed NDVI1, was optimal for biomass as there was difficulty computing the covariance matrix for two and three covariates in the grid and triangular schemes. Cokriging with the log transformed wide IR and log transformed 660 nm wavebands was optimal for cokriging with percent grass cover. Three covariates, log transformed 660 nm, 460 nm and the log transformed wide IR band, were optimal for percent legume cover. Covariate selection was determined based upon correlation with the plant parameter of interest (Table 11). The minimization of kriging RMSE for cross-validation sets aided in determining which combination of covariates was optimal.

Map Results

Kriging vs. Cokriging

In general, a visual variance existed between the kriged and cokriged maps. The cokriged maps exhibited more short-range variation and local detail in their depiction of the variability of the three plant parameters. However, both methods resulted in similar patterns of variability for the plant parameters. Comparable results have been found in the soil literature for topsoil silt cokriged with subsoil silt and sand as covariates (McBratney and Webster, 1983a), $NaHCO_3$-extractable P cokriged with 25 percent HCl-extractable P

(Trangmar et al., 1986), and NO_3 cokriged with soil EC (Zhang et al., 1992). Improved local detail of the cokriging maps was due to the finer sampling grid of the covariate(s), canopy reflectance (McBratney and Webster, 1983a). Representative maps of these observations are shown in Figure 15. The actual differences resulting from these visual variances will be analyzed quantitatively.

Sampling scheme comparison

Based on the design and results of this experiment, it cannot be concluded which of the three sampling schemes is optimal for producing accurate maps of variability for vegetative characteristics. Each of the schemes used in this study was just one depiction of a grid scheme or a triangular scheme or a random scheme in the pasture. Multiple scenarios of each scheme should be analyzed further in order to answer this question. General conclusions about the sampling schemes will be discussed after examining them quantitatively.

Quantitative Results

Kriging vs. Cokriging

When comparing results estimated with kriging and cokriging, two parameters were analyzed. Root mean square error of prediction and the correlation between predicted values from kriging and the actual values taken from direct plant measurements were evaluated. As previously mentioned, validation sets of 86 sample points were available for the n = 30 sampling schemes. The RMSE should be small for an unbiased and precise prediction. For both kriging and cokriging and for each sampling scheme, the root mean square error was calculated as:

$$\mathrm{RMSE} = \left\{ \frac{1}{n} \sum_{i=1}^{n} [z(s_i) - z^*(s_i)]^2 \right\}^{0.5}$$

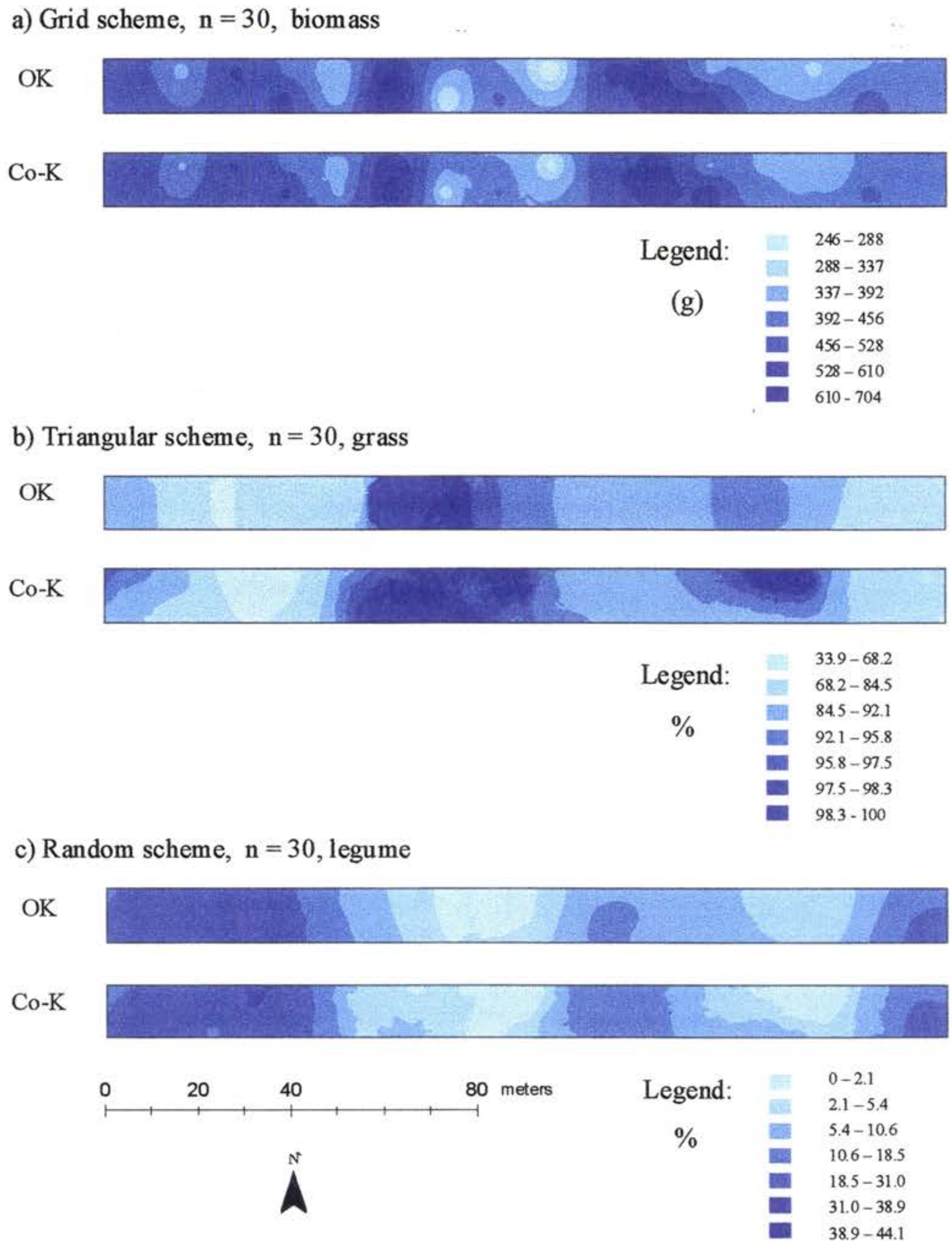

Figure 15. Representative comparisons of kriged (OK) and cokriged (Co-K) maps for a) n = 30 grid sampling scheme, biomass, b) n = 30 triangular sampling scheme, % grass and c) n = 30 random sampling scheme, % legume.

where n is the number of sample sites in the validation set, $z(s_i)$ are the observed plant values and $z^*(s_i)$ are the predicted values. As shown in Table 12, RMSE for cokriging was consistently lower than the RMSE for kriging, with one exception. Thus, cokriging helped improve the prediction of the validation sites in all the scenarios but one. The single exception occurred for biomass sampled with the random scheme. This result may be explained by the large nugget variance of biomass. In addition, the correlation of biomass with its covariate, NDVI1 ($r = 0.43$), was not remarkably significant. Furthermore, a random sampling scheme often has "gaps" of unsampled space, so the prediction of biomass values was more reliant upon the variability in neighboring covariate points than farther away biomass points.

Correlation between predicted and actual values increased for most of the plant variables and sampling schemes when cokriging was implemented (Table 12). Based on the increase in RMSE with cokriging biomass in the random scheme, the decrease in correlation between predicted and actual values was not surprising. The correlation between predicted and actual values also decreased with cokriging biomass in the triangular scheme. However, the very low correlation between predicted and actual values for kriging biomass in the triangular scheme indicated that the spatial autocorrelation was difficult to model. When the target variable has a high nugget variance, gains from cokriging are not likely (Webster and Oliver, 2001).

Figure 16 illustrates the reduction in RMSE of prediction due to cokriging as a function of the absolute correlation between the reflectance covariates and the three target plant parameters. Relative reduction in RMSE was defined by:

$$100(RMSE_k - RMSE_{ck})/RMSE_k$$

where $RMSE_k$ and $RMSE_{ck}$ are the root mean square errors of kriging and cokriging, respectively (adapted from Zhang et al., 1992). Several important observations can be made

Table 12. Validation set root mean square errors (RMSE) of plant data prediction for kriging and cokriging the three sampling schemes. Coefficient of determination (r^2) between predicted and actual values for validation set.

Target variable	Sample size		Grid		Triangular		Random	
			OK†	Co-K	OK	Co-K	OK	Co-K
Biomass, g	30	RMSE‡	81.78	76.8 (-)§	93.2	89.45 (-)	91.96	95.24 (+)
		r^2	0.2336	0.2608 (+)	0.0772	0.0559 (-)	0.0676	0.0165 (-)
Grass, %	30	RMSE	14.11	12.97 (-)	13.25	12.99 (-)	14.18	14.11 (-)
		r^2	0.3158	0.4257 (+)	0.3431	0.3651 (+)	0.3023	0.3143 (+)
Legume, %	30	RMSE	13.87	12.82 (-)	12.78	12.7 (-)	13.35	12.8 (-)
		r^2	0.3171	0.4232 (+)	0.3712	0.3775 (+)	0.3854	0.4298 (+)

† OK is ordinary kriging, Co-K is cokriging.

‡ RMSE units are specific to each plant variable.

§ Plus or minus sign in parentheses indicates an increase (+) or decrease (-) in RMSE or r^2 value between OK and Co-K.

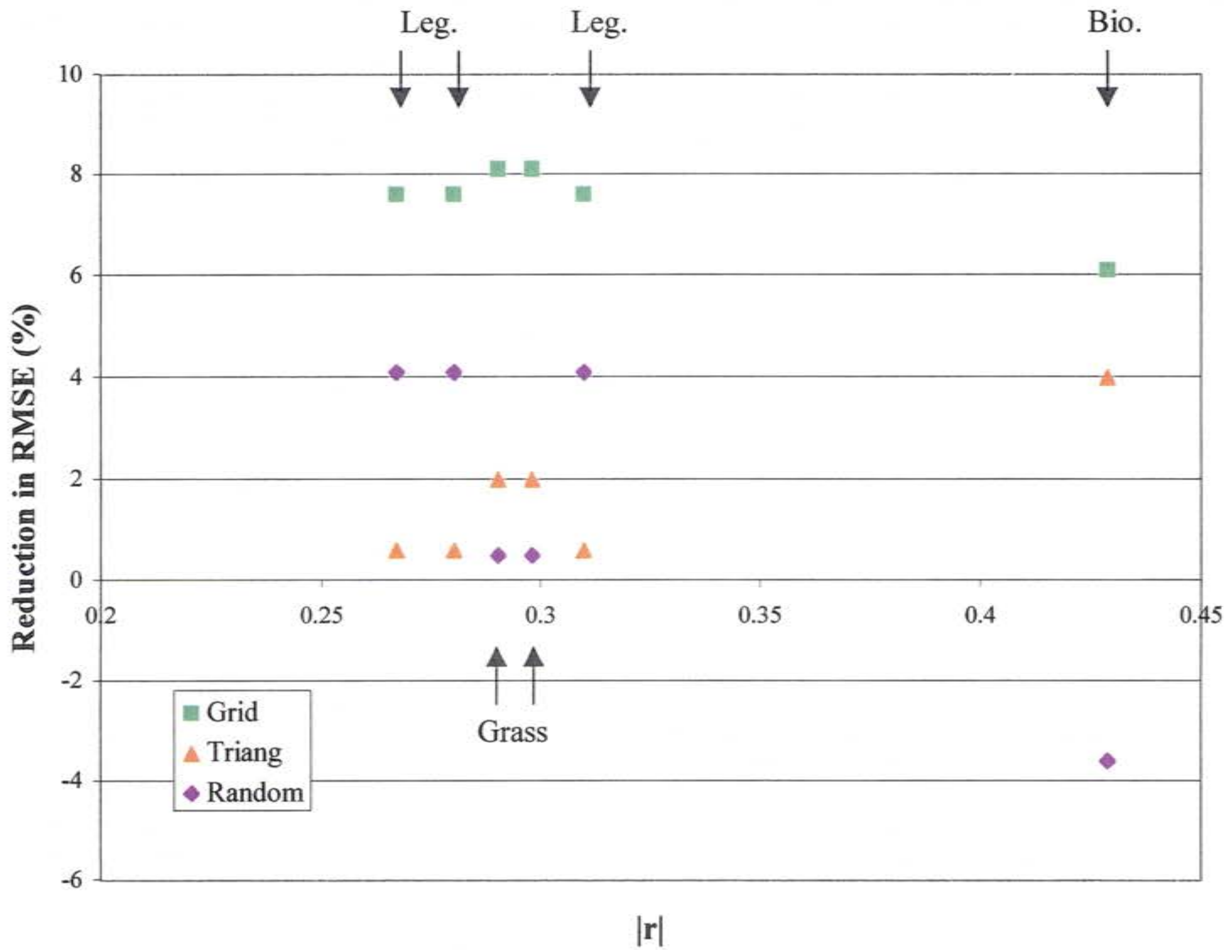

Figure 16. Reduction in root mean square error (RMSE) of prediction due to cokriging as a function of the absolute correlation (r) between canopy reflectance at covariate wavebands and the three target plant variables: biomass (Bio.), % grass (Grass) and % legume (Leg.). Data is plotted for each sampling pattern.

from this figure. First, a higher correlation between the target plant variable and canopy reflectance wavebands did not consistently improve reduction in RMSE. Yates and Warrick (1987) found that a reduction in kriging variance was observed as the correlation between the target variable and covariate increased. However, in this study, two covariates were used to predict percent grass cover and three covariates were used to predict percent legume cover. For the grid and the random sampling schemes, it appears that use of multiple covariates resulted in larger reductions in RMSE. This result concurred with McBratney and Webster's observation that using two covariates resulted in more precise cokriging estimations than a single covariate (1983a).

General trends can be drawn from Figure 16, but it is important to note that the efficacy of cokriging is more than a function of correlation between a covariate and target variable. The efficacy also includes the strength of the spatial cross-correlation between the covariate and target variable, the geometric sampling pattern and the ratio of the sampling intensities of the covariate and target variable (McBratney and Webster, 1983a).

Sampling scheme comparison

The mapping accuracy from cokriging was also affected by the sampling pattern used. Also from Figure 16, it is evident that the largest reductions in RMSE due to cokriging were found with the grid sampling scheme. This result is likely due to the more systematic, geometric sampling of the grid scheme. Both Vauclin et al. (1983) and McBratney and Webster (1983a) found that cokriging consistently reduced estimation variances where target and covariate properties were sampled in geometric patterns. However, because only one example of each sampling scheme was used, generalizations about each of the schemes cannot be made.

Conclusions

Use of a surrogate measure such as plant canopy reflectance can be beneficial in predicting unsampled areas of a pasture. Maps resulting from cokriging reflectance values with biomass, percent grass cover and percent legume cover exhibited more local detail than the kriged maps of each plant parameter. The use of canopy reflectance as a covariate improved prediction of grass and legume percent cover in all three sampling schemes studied. The prediction of above ground biomass was not quite as consistent; however, this was likely due to the low amount of spatial continuity of biomass values.

This study showed an overall improvement in kriging variance of unsampled sites when cokriging was implemented. The grid sampling scheme appeared to benefit most from cokriging, but the results are inconclusive because only one grid scheme (not multiple) was analyzed.

Use of rapid and indirect methods for quantifying pasture variability could provide useful and convenient information for more precise pasture management. Ground-based radiometers provide a rapidly, densely, noninvasively and easily collected method for quantifying plant variability. The use of NDVI as a covariate improved estimation accuracy of pasture biomass when a grid sampling scheme was used. This result could insure more accurate estimates of pasture-wide productivity. Species composition is an indication of pasture quality. This study showed that using waveband spectra as covariates can also lead to more accurate estimation of pasture composition. Thus, use of spectral reflectance as a covariate indirectly improved estimation of pasture quantity and quality.

References

Aquino, V.M., F.M. Shokes, R.D. Berger, D.W. Gorbet, and T.A. Kucharek. 1992. Relationships among late leafspot, healthy leaf area duration, canopy reflectance, and pod yield of peanut. Phytopathology 82:546-552.

Blackmer, T.M., J.S. Schepers, G.E. Varvel, and E.A. Walter-Shea. 1996. Nitrogen deficiency detection using reflected shortwave radiation from irrigated corn canopies. Agron. J. 88:1-5.

Bouman, B.A.M., D. Uenk, and A.J. Haverkort. 1992. The estimation of ground cover of potato by reflectance measurements. Potato Research 35:111-125.

Burgess, T.M., and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil Sci. 31:315-331.

Chien, Y., D. Lee, H. Guo, and K. Houng. 1997. Geostatistical analysis of soil properties of mid-west Taiwan soils. Soil Science 162:291-298.

Daubenmire, R. 1959. A canopy-coverage method of vegetational analysis. Northwest Science 33:43-64.

ESRI. 1996. ArcView Spatial Analyst, Redlands, CA.

ESRI. 2001. ArcView 8.1. Release ArcGIS Geostatistical Analyst, Redlands, CA.

Guan, J., and F.W. Nutter, Jr. 2001. Factors affecting the quality and quantity of sunlight reflected from alfalfa canopies. Plant Dis. 85:865-874.

Heisel, T., A.K. Ersboll, and C. Andreasen. 1999. Weed mapping with co-kriging using soil properties. Precision Agriculture 1:39-52.

Journel, A.G., and C.J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York.

Kravchenko, A.N., and D.G. Bullock. 2002. Spatial variability of soybean quality data as a function of field topography: I. Spatial Data Analysis. Crop Science 42:804-815.

Lawrence, P.R., B. Gerard, C. Moreau, F. Lheriteau, and A. Buerkert. 2000. Design and testing of a global positioning system-based radiometer for precision mapping of pearl millet total dry matter in the Sahel. Agron. J. 92:1086-1095.

Ma, B.L., M.J. Morrison, and L.M. Dwyer. 1996. Canopy light reflectance and field greenness to assess nitrogen fertilization and yield of maize. Agron. J. 88:915-920.

Matheron, G. 1971. The theory of regionalized variables and its applications, Vol. Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau, No. 5. Ecole Nationale Superieure des Mines de Paris.

McBratney, A.B., and R. Webster. 1983a. Optimal interpolation and isarithmic mapping of soil properties. V. Co-regionalization and multiple sampling strategy. Journal of Soil Science 34:137-162.

McBratney, A.B., and R. Webster. 1983b. How many observations are needed for regional estimation of soil properties? Soil Science 135:177-183.

Mitchell, A.R., P.J. Pinter, Jr., J.N. Guerrero, C.B. Hernandez, and V.L. Marble. 1990. Spectral reflectance measurements of alfalfa under sheep grazing. Agron. J. 82:1098-1103.

Nilsson, H., and L. Johnsson. 1996. Hand-held radiometry of barley infected by barley stripe disease in a field experiment. Journal of Plant Diseases and Protection 103:517-526.

Nutter, F.W., Jr. 1989. Detection and measurement of plant disease gradients in peanut with a multispectral radiometer. Phytopathology 79:958-963.

Nutter, F.W., Jr., and R.H. Littrell. 1996. Relationships between defoliation, canopy reflectance and pod yield in the peanut-late leafspot pathosystem. Crop Protection 15:135-142.

Nutter, F.W., Jr., R.H. Littrell, and T.B. Brenneman. 1990. Utilization of a multispectral radiometer to evaluate fungicide efficacy to control late leaf spot in peanut. Phytopathology 80:102-108.

Oliver, M.A. 1987. Geostatistics and its application to soil science. Soil Use and Management 3:8-20.

Olson, K.C., and R.C. Cochran. 1998. Radiometry for predicting tallgrass prairie biomass using regression and neural models. J. Range Management 51:186-192.

Trangmar, B.B., R.S. Yost, and G. Uehara. 1986. Spatial dependence and interpolation of soil properties in West Sumatra, Indonesia: II. Co-regionalization and co-kriging. Soil Sci. Soc. Am. J. 50:1396-1400.

Triantafilis, J., I.O.A. Odeh, and A.B. McBratney. 2001. Five geostatistical models to predict soil salinity from electromagnetic induction data across irrigated cotton. Soil Sci. Soc. Am. J. 65:869-878.

Vauclin, M., S.R. Vieira, G. Vachaud, and D.R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175-184.

Ver Hoef, J.M. 2002. Personal communication. Alaska Department of Fish and Game, Fairbanks, Alaska.

Vrindts, E., J. De Baerdemaeker, and H. Ramon. 2002. Weed detection using canopy reflection. Precision Agriculture 3:63-80.

Walburg, G., M.E. Bauer, C.S.T. Daughtry, and T.L. Housley. 1982. Effects of nitrogen nutrition on the growth, yield, and reflectance characteristics of corn canopies. Agron. J. 74:677-683.

Warrick, A.W., D.E. Myers, and D.R. Nielsen. 1986. Geostatistical methods applied to soil science, *In* A. Klute, ed. Methods of soil analysis, Part 1. 2nd ed. Agronomy 9:53-82.

Webster, R., and M.A. Oliver. 2001. Geostatistics for environmental scientists. John Wiley & Sons, Ltd., Chichester.

Webster, R., P.J. Curran, and J.W. Munden. 1989. Spatial correlation in reflected radiation from the ground and its implications for sampling and mapping by ground-based radiometry. Remote Sensing of Environment 29:67-78.

Yates, S.R., and A.W. Warrick. 1987. Estimating soil water content using cokriging. Soil Sci. Soc. Am. J. 51.

Zhang, R., D.E. Myers, and A.W. Warrick. 1992. Estimation of the spatial distribution of soil chemicals using pseudo-cross-variograms. Soil Sci. Soc. Am. J. 56:1444-1452.

CHAPTER 6. CONCLUSIONS

The measurement of soil electroconductivity (EC) by electromagnetic induction is a valuable tool for quantifying variation in soil properties related to moisture and texture in a central Iowa pasture. With respect to the first objective, variation in soil EC provided direction for development of a stratified sampling scheme. A multistage sampling scheme incorporated two types of noninvasively collected data: soil EC and topographic measures of the pasture. Delineation of the pasture into five zones based on a fuzzy k-means classification algorithm resulted in relatively homogeneous sampling zones. While the zones were not unique among all five soil variables of interest, they were created with fairly little effort and were based on noninvasive measures. The zones served as groundwork for development of a multistage sampling scheme.

Compared to random sampling schemes of n = 30 and n = 15, the multistage sampling schemes of equal size reduced the variance of the estimated population total for four of the five soil variables. The soil variables most closely related to those used for classification were predicted with the least error at unsampled points. In general, there was a loss in prediction accuracy resulting from a decrease in sampling intensity from n = 30 to n = 15. Stratification using noninvasively collected data, a fuzzy k-means classification algorithm and ranked set sampling (McIntyre, 1952) resulted in a more directed approach to soil sampling.

Coregionalization among soil EC and five soil variables of interest was explored through the geostatistical method of cokriging. Relative to the second objective of the study, maps of the five soil variables cokriged with soil EC exhibited increased local detail in variation. Exceptions occurred for the soil variables displaying short-range spatial variability. Across all five target variables, four sampling schemes and two sampling densities, an overall, but inconsistent, improvement in kriging variance and prediction accuracy of unsampled sites was received when soil variables were cokriged with soil EC.

The improvement with cokriging was generally greater for soil variables more highly correlated with soil EC.

When examining the two sampling densities, similar spatial patterns were evident in the maps. However, a clear improvement from cokriging with a lower density target variable was not observed when quantitative comparisons of prediction accuracy were made. From the sampling scheme comparisons, the triangular grid appeared to benefit most from cokriging, but the results were inconclusive because only one triangular scheme (not multiple) was analyzed. Integrating the dense base of information easily and rapidly provided by soil EC with less densely and invasively collected soil parameters contributed to more accurate maps of pasture soil variability.

Using a ground-based radiometer, certain waveband spectra were found to be more highly correlated with above ground biomass, percent grass cover and percent legume cover than others. Relative to objective three, biomass was most highly correlated with NDVI1, percent grass cover with reflectance at 660 nm and at the wide IR band and percent legume cover with reflectance at 460 nm, 660 nm and at the wide IR band. These relationships were capitalized upon by cokriging in the fourth objective of this study. Using one, two and three covariates, the cokriged maps exhibited more short-range variation and local detail in their depiction of the variability of biomass, percent grass cover and percent legume cover, respectively. The use of canopy reflectance as a covariate improved prediction of grass and legume percent cover in all three sampling schemes studied. The prediction of above ground biomass was not quite as consistent. Canopy reflectance was an easily and rapidly collected surrogate measure that provided valuable information used in predicting unsampled areas of a pasture.

Several implications resulted from this study. Because soil EC and canopy reflectance are measured in such a rapid, easy, noninvasive and inexpensive manner, collection of dense data may allow for fewer points of invasive sampling. Knowing about field variation without an invasive and time-consuming survey may save labor and can direct efforts for sampling. Secondly, the coupling of indirect sampling methods with GPS adds an important spatial dimension to pasture characterization. Spatial data can tell a story by associating location with every data point. In addition, with improved characterization of pasture variability comes improved management. With respect to soil, improved characterization could lead to more accurate maps and more precise management of field inputs such as nutrients and lime. On the plant side, use of spectral reflectance as a covariate could indirectly improve estimation of pasture quantity, in terms of above ground biomass, and quality, in terms of estimating percent cover of grass and legume.

ACKNOWLEDGMENTS

In my years long before graduate school, my mom used to encourage me to write thank you notes to family and friends for all my birthday and Christmas gifts. One might say that the good habit stuck, and I'd like to continue that habit.

Starting from my beginning, I would like to thank my wonderful parents. My mom and dad encouraged me to go wherever, do whatever, but always use the talents that were given to me. I am so grateful for them, and I know my happiness makes them happy.

Many thank you's to my committee members. Dr. Dixon, I really appreciated your statistical expertise. You helped me improve my statistical knowledge. Lee, thank you for your soil morphology expertise. Yes, I do see the beauty in soils. Mary, thank you for your advice and encouragement. You have allowed me to experience more than just research while in graduate school.

I appreciated working with Dr. Fenton and sharing in his soil EC wisdom. Thank you for the use of your EM-38 and for helping to identify soil characteristics at my pasture site. Eric Brevik was also very influential in expanding my knowledge of soil EC. Thank you to Eric for adding to my literature collection. Thanks to Don Bullock for sampling scheme advice and to Susan White and Tobey Ewing for technical support in the planning stages of my research.

To my 1541 officemates, Carla Rogis and Lori Hinze, who were with me through it all – thanks a bunch!

My research could not have been done without a great team. I am very grateful to have had Amy Osterhaus, Julie Wheelock, Jenny Carpenter, and Carmen Wiltse on my team. Their hard work and great personalities made for a fun work environment. Also included on this forage team were Indi Braden, John Guretzky, Bryan Terrell, Todd White, Keith Betteridge, Roger Hintz, Trish Patrick and Ken Moore. What a great couple of seasons we had together.

Indi and John, thank you for your GPS and GIS expertise (and sharing similar frustrations!). Indi, may your batteries always stay charged and remember: in cases of GPS technology, two heads are usually better than one! John, thank you for always noticing things and for sharing your pastures with me. Bryan, thank you for accommodating my computer preferences. Todd, you always provided comic relief – thanks for your Kiwi humor and pavlov(er).

Roger and Trish, thank you so much for your incredible help and support not only during my research but all other times. Because of you both, my research ran like clockwork.

Roger, I now see that the sun always shines in Iowa and thanks for all the "shop talk" along with the style-ish humor. Trish, thank you for your super organization, being both a great listener and advice-giver and for the pop credit system.

I believe that life is what you make of it, but I'd like to thank Ken for helping me to make more out of it. Ken, I learned a great deal from your example, and I hope to carry that with me in both my personal and professional career. Thank you for suggesting advancement opportunities, always making time for me, and allowing me to have really great technology and resources readily available.

And finally, to my husband Gabe, thank you for being you. You supplied the humor when I needed it most. Thank you for being so understanding when I came home late or worked in the office night after night. Your constant support and reassurance helped me through things. You are great to swap ideas with. However we managed it, thank you for helping me balance all that wedding planning with everything else. Oh, and thank you for all your help on my thesis.